The Dissertation Committee for Raghuprasad Sidharthan certifies that this is the approved version of the following dissertation:

On the Estimation and Application of Flexible Unordered Spatial Discrete Choice Models

Committee:

Chandra R. Bhat, Supervisor

Jason Abrevaya

Steve Boyles

Randy B. Machemehl

Michael Oden

Chandler Stolp

On the Estimation and Application of Flexible Unordered Spatial Discrete Choice Models

by

Raghuprasad Sidharthan, B.Tech.; M.Tech.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2012

Dedication

To my late grandfather V. Karunakaran Nair.

To my father, Sidharthan Thottathil, and my mother, Parvathi Veluthakkal, for supporting my decisions throughout.

To all my teachers for inspiring me.

Acknowledgements

First, I would like to express my sincere gratitude to my advisor, Dr. Chandra Bhat, for his continuous guidance, help and encouragement throughout my doctoral studies. In addition to that I am grateful to all members of my dissertation committee, Dr. Jason Abrevaya, Dr. Steve Boyles, Dr. Randy B. Machemehl, Dr. Michael Oden, and Dr. Chandler Stolp for providing helpful comments and suggestions throughout the course my dissertation research. I thank Dr. Ram Pendyala for the research feedbacks he has given. I thank Lisa Macias for being always helpful and quick in assisting in all matters. I also thank Danny Quiroz for helping with computing resources and for being so approachable.

My PhD journey would not have been as much enjoyable without the company and support of my fellow doctoral students Rajesh Paleti and Marisol Castro. The intellectual discussions we had, advice and encouragement they gave me will always be remembered. I also want to thank Binny Paul, my roommate for two years, for being such a good friend. The racquetball sessions that I had with Binny, Rajesh and our other friend Roshan will always be cherished. Finally I would like to thank my soccer friends from the Clarks field for making my Wednesday evenings thoroughly enjoyable.

Last but not least I would like to thank my mother, Parvathi, my father, Sidharthan, my sister, Radha, my grandmother Sridevi and my brother-in-law Vinod for the unconditional love and support. I would also like to thank my uncles, aunts and cousins for their love.

On the Estimation and Application of Flexible Unordered Spatial Discrete Choice Models

Raghuprasad Sidharthan, Ph.D.

The University of Texas at Austin, 2012

Supervisor: Chandra R. Bhat

Unordered choice models are commonly used in the field of transportation and several other fields to analyze discrete choice behavior. In the past decade, there have been substantial advances in specifying and estimating such models to allow unobserved taste variations and flexible error covariance structures. However, the current estimation methods are still computationally intensive and often break down when spatial dependence structures are introduced (due to the resulting high dimensionality of integration in the likelihood function). But a recently proposed method, the Maximum Approximate Composite Marginal Likelihood (MACML) method, offers an effective approach to estimate such models. The MACML approach combines a composite marginal likelihood (CML) estimation approach with an approximation method to evaluate the multivariate standard normal cumulative distribution (MVNCD) function. The composite likelihood approach replaces the likelihood function with a surrogate likelihood function of substantially lower dimensionality, which is then subsequently evaluated using an analytic approximation method rather than simulation techniques. This combination of the CML with the specific analytic approximation for the MVNCD function is effective because it involves only univariate and bivariate cumulative normal distribution function evaluations, regardless of the dimensionality of the problem.

For my dissertation, I have four objectives. The first is to evaluate the performance of the MACML method to estimate unordered response models by undertaking a Monte Carlo simulation exercise. The second is to formulate and estimate a spatial and temporal unordered discrete choice model and apply this model to a land use change context and to the mode choice decision of school children. The third objective is to formulate a random coefficient model with non-normal mixing distributions on model parameters which can be estimated using the MACML approach. Finally, the fourth objective us to propose an improvement to the MACML method by incorporating a second order MVNCD function that is more accurate and evaluate its performance in estimating parameters for a variety of model structures.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND: EVOLUTION OF DISCRETE CHOICE MODELS

Unordered choice models are commonly used in the field of transportation and several other fields to analyze discrete choice behavior. In the past decade, substantial advances have been made in specifying and estimating such models to allow unobserved taste variations (see Bhat and Sardesai, 2006) and flexible error covariance structures (see Train, 2009, Chapter 5). A common approach in these flexible models is to superimpose a normal mixing distribution over a generalized extreme-value (GEV) kernel. Such a normally mixed GEV model structure has been preferred over a multinomial probit (MNP) model structure because of the relative ease with which a mixed GEV model can be estimated using the maximum simulated likelihood (MSL) inference approach (see Bhat (2003) and Bhat *et al*., (2008)). Despite this widespread use of the MSL estimation method, the consistency, efficiency, and asymptotic normality of the MSL estimator critically depends on the number of simulation draws used. Due to computational cost involved in using very high number of simulation draws, a lower number of draws is often used, leading to a reduction in the estimation efficiency and causing convergence issues. Additionally, an issue that is usually not considered is the relatively poor accuracy of the covariance matrix (obtained from the hessian of the MSL function) due to the highly non-linear nature of the MSL function. All these considerations become particularly problematic as the dimensions of the integration increases. However, a recently proposed method by Bhat (2011a), the Maximum approximate composite marginal likelihood (MACML) method, offers an effective approach to estimate unordered choice models with flexible error structures and taste variation.

In the past decade, in addition to relaxing the assumptions on the error term and the inter individual taste variations, there has been increasing attention in discrete choice modeling on accommodating spatial dependence across decision agents or observational units to recognize the potential presence of diffusion effects, social interaction effects, or unobserved location-related influences (see Jones and Bullen, 1994, and Miller, 1999). Specifically, spatial lag and spatial error-type structures developed in the context of

continuous dependent variables to accommodate spatial dependence (see, for instance, Dubin, 1998, Cho and Rudolph, 2007, Anselin, 2006, Elhorst, 2010ab) are being considered for discrete choice dependent variables (see reviews of this literature in Franzese *et al*. 2010, Brady and Irwin, 2011, and Bhat *et al*., 2010a). But almost all of this research focuses on binary or ordered response choice variables by applying global spatial structures to the linear (latent) propensity variables underlying the choice variables (for example, see Fleming, 2004, Franzese and Hays, 2008, Franzese *et al*., 2010, and LeSage and Pace, 2009). The two dominant techniques, both based on simulation methods, for the estimation of such spatial binary/ordered discrete models are the frequentist recursive importance sampling (RIS) estimator (which is a generalization of the more familiar Geweke-Hajivassiliou-Keane or GHK simulator; see Beron and Vijverberg, 2004) and the Bayesian Markov Chain Monte Carlo (MCMC)-based estimator (see LeSage and Pace, 2009). However, both of these methods are confronted with multi-dimensional normal integration, and are cumbersome to implement in typical empirical contexts with moderate to large estimation sample sizes (see Bhat, 2011a and Smirnov, 2010).

The RIS and MCMC methods become even more difficult to implement in a spatial unordered multinomial choice context because the likelihood function entails a multidimensional integral of the order of the number of observational units factored up by the number of alternatives minus one (in the case of multi-period data the integral dimension gets factored up further by the number of time periods of observation). Thus, it is no surprise that there has been little research on including spatial dependency effects in unordered choice models. However, here also the MACML method can be used in the estimation, and requires no simulation. The MACML estimation of spatial MNP models involves only univariate and bivariate cumulative normal distribution function evaluations, regardless of the number of alternatives or the number of choice occasions per observation unit, or the number of observation units, or the nature of social/spatial dependence structures. The MACML method is based on the composite marginal

likelihood (CML) method and therefore we will discuss the CML method briefly before providing the details for the MACML method.

1.2 THE COMPOSITE MARGINAL LIKELIHOOD APPROACH

The CML estimators are used for problems where the maximum likelihood estimations are computationally time consuming (or nearly infeasible) due to the complexity and/or dimensionality of the problem. The CML method belongs to a class of more general methods called as the composite likelihood methods proposed by Lindsay (1988). Varin *et al*., 2011 provides a comprehensive and recent review of the CML approaches. The CML approach represents a conceptually and pedagogically simple simulation-free procedure relative to simulation techniques. The approach may be explained in a simple manner by taking the case of modeling random coefficient (time invariant coefficients) for a panel or repeated choice data. In such a case, instead of considering the likelihood of all the time periods, the CML method considers a surrogate likelihood function that is the product of the probability of easily computed marginal events. For instance, one may compound (multiply) pairwise probabilities of the choice of individual q at time t *and* at time t', of the choice of individual q at time t *and* at time t'', and so on and so forth. The CML estimator is then the one that maximizes the compounded probability of all pairwise events (see Varin and Vidoni, 2009, Engle *et al*., 2007, Bhat *et al*., 2010b, and Bhat and Sener, 2009 for applications of the estimator for binary and ordered-response systems). Alternatively, the analyst can also consider larger subsets of observations, such as triplets or quadruplets or even higher dimensional subsets (see Engler *et al*., 2006 and Caragea and Smith, 2007). However, doing so defeats the purpose of the approach because it leads to high dimensionality of integration. Besides, it is generally agreed that the pairwise approach is a good balance between statistical and computational efficiency.

The properties of the general CML estimator may be derived using the theory of estimating equations (see Cox and Reid, 2004). Specifically, under usual regularity assumptions (Molenberghs and Verbeke, 2005, page 191), the CML estimator is consistent and asymptotically normal distributed (this is because of the unbiasedness of

the CML score function, which is a linear combination of proper score functions associated with the marginal event probabilities forming the composite likelihood). Consider a parameter vector $\boldsymbol{\theta}$ being estimated using the CML estimator and let $\hat{\boldsymbol{\theta}}_{CML}$ be the CML estimate. Now the asymptotic normality can be written as follows:

$$\sqrt{n}\,(\hat{\boldsymbol{\theta}}_{CML} - \boldsymbol{\theta}) \sim MVN(0, V_{CML}(\hat{\boldsymbol{\theta}}_{CML})) \tag{1}$$

Where n is the sample size and $MVN(\cdot)$ is the multivariate normal distribution and $V_{CML}(\hat{\boldsymbol{\theta}}_{CML})$ is the Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005) given by:

$$V_{CML}(\hat{\boldsymbol{\theta}}_{CML}) = [G(\boldsymbol{\theta})]^{-1} = [H(\boldsymbol{\theta})]^{-1} J(\boldsymbol{\theta}) [H(\boldsymbol{\theta})]^{-1}, \text{ where} \tag{2}$$

$$H(\boldsymbol{\theta}) = E\left[-\frac{\partial^2 \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] \text{ and}$$

$$J(\boldsymbol{\theta}) = E\left[\left(\frac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \right) \right].$$

where $L_{CML}(\cdot)$ is the composite marginal likelihood. The "bread" matrix $H(\boldsymbol{\theta})$ of Equation (2) can be estimated in a straightforward manner using the Hessian of the negative of the CML likelihood function, evaluated at the CML estimate $\hat{\boldsymbol{\theta}}_{CML}$. The "vegetable" matrix $J(\boldsymbol{\theta})$ is not that straightforward to estimate and depends on the dependency structure of the model. For a case of panel data model where the dependency is only across time periods of the same individual and not across individuals, a clustering method can be used. On the other hand, for a case of spatial model in which there is global interdependence a different method such as the windows sampling method proposed by Heagerty and Lumley (2000) is required (see Chapter 4 for details).

Many of the model selection criterion developed for the maximum likelihood estimation approaches have been extended to the CML estimators. Simple t-statistic test based on the Godambe's matrix can be used to retain or reject a single parameter. When there are multiple parameters involved in the test (nested models), then the composite

likelihood ratio test (CLRT) statistic can be used. Consider the null hypothesis $H_0 : \boldsymbol{\tau} = \boldsymbol{\tau_0}$ against $H_1 : \boldsymbol{\tau} \neq \boldsymbol{\tau_0}$, where τ is a subvector of $\boldsymbol{\theta}$ of dimension *d*; *i.e.*, $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\alpha}')'$. Then CLRT is given by:

$$CLRT = 2[\log L_{CML}(\hat{\boldsymbol{\theta}}) - \log L_{CML}(\hat{\boldsymbol{\theta}}_{\mathbf{0}})], \tag{3}$$

where $\hat{\boldsymbol{\theta}}$ is the CML estimator of the unrestricted model, and $\boldsymbol{\theta_0}$ is the CML estimator for the restricted model (and nested). However, the distribution of this CLRT statistic is not a standard chi-squared asymptotic distribution and therefore to use it in model selection one has to do parametric bootstrapping to obtain its distribution, which makes it cumbersome. As an alternative to the CLRT statistic Pace *et al.* (2011) have recently proposed the adjusted composite likelihood ratio test (AD*CLRT*) statistic which allows for inference based on the asymptotic chi-square distribution (with *d* degrees of freedom) and is given by the following expression:

$$ADCLRT = \frac{[\boldsymbol{S}_{\tau}(\boldsymbol{\theta})]'[\boldsymbol{H}_{\tau}(\boldsymbol{\theta})]^{-1}[\boldsymbol{G}_{\tau}(\boldsymbol{\theta})][\boldsymbol{H}_{\tau}(\boldsymbol{\theta})]^{-1}\boldsymbol{S}_{\tau}(\boldsymbol{\theta})}{[\boldsymbol{S}_{\tau}(\boldsymbol{\theta})]'[\boldsymbol{H}_{\tau}(\boldsymbol{\theta})]^{-1}\boldsymbol{S}_{\tau}(\boldsymbol{\theta})} \times CLRT \tag{4}$$

where $\boldsymbol{S}_{\tau}(\boldsymbol{\theta})$ is the $d \times 1$ submatrix of $\boldsymbol{S}(\boldsymbol{\theta}) = \left(\frac{\partial \log L_{CML}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)$ corresponding to the vector $\boldsymbol{\tau}$, and all the matrices above are computed at $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$.Finally, when two non-nested models are to be compared and selected, the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (2005) which is a generalized form of the familiar Akaike's Information Criterion (AIC) may be used. This statistic is given by:

$$\log L^{*}_{CML}(\hat{\boldsymbol{\theta}}) = \log L_{CML}(\hat{\boldsymbol{\theta}}) - tr\left[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{H}}(\hat{\boldsymbol{\theta}})^{-1}\right] \tag{5}$$

The model that provides a higher value of CLIC is selected.

1.3 THE MAXIMUM APPROXIMATE COMPOSITE LIKELIHOOD APPROACH

The MACML method is an extension to the CML approach and is applicable to broader classes of problem. This is achieved by using an approximation method to evaluate the

multivariate standard normal cumulative distribution (MVNCD) function whenever the pairwise compounded probability of the CML function requires the computation of a multivariate CDF of order greater than two. This extension widens the applicability of the CML method to a wider range of problems including unordered choice models. In contrast to the approaches that are based on evaluating the multidimensional integrals in the true likelihood function using simulation techniques, the MACML estimation approach for cross-sectional unordered-response models with normally distributed mixing is based on analytic approximations to the MVNCD functions in the true likelihood function. While the approximation used in the MACML method may not be very accurate at an individual probability level, it works well in the estimation because of the large number of observations involved in the evaluation of the sample likelihood. The approximation is based on the decomposition of the multivariate cumulative normal probability into a product of conditional probabilities. The approximation can be described by considering a normal random vector $\boldsymbol{W}$ of dimension I $(W_1, W_2, W_3, \ldots, W_I)$ with mean zero and variance 1 for each of the dimensions. The correlation matrix is assumed to be $\boldsymbol{\Sigma}$. Now, the orthant probability that we intend to evaluate can be written as:

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1,\ W_2 < w_2,\ W_3 < w_3, \ldots,\ W_I < w_I) \tag{6}$$

This can be written as a product of conditional probabilities as given below for $I \geq 3$:

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1,\ W_2 < w_2) \times \prod_{i=3}^{I} \Pr(W_i < w_i \mid W_1 < w_1,\ W_2 < w_2,\ W_3 < w_3, \ldots,\ W_{i-1} < w_{i-1}). \tag{7}$$

Next, we define the indicator variable $\tilde{I}_i = I(W_i < w_i)$, where $I(\cdot)$ is the indicator function that takes a values 1 if the condition within the parenthesis is true and zero otherwise. This also implies that $E(\tilde{I}_i) = \Phi(w_i)$, where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a univariate normal distribution. Using these notations, equation (7) can be written as

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1,\ W_2 < w_2) \times \prod_{i=3}^{I} E\,(\tilde{I}_i \mid \tilde{I}_1 = 1,\ \tilde{I}_2 = 1, \tilde{I}_3 = 1, ...,\ \tilde{I}_{i-1} = 1). \qquad (8)$$

In equation (8) above, an exact computation is performed for the first term, which is a bivariate CDF, and the following approximation is used for the conditional probabilities:

$$\begin{aligned} & E\,(\tilde{I}_i \mid \tilde{I}_1 = 1,\ \tilde{I}_2 = 1, \tilde{I}_3 = 1, ...,\ \tilde{I}_{i-1} = 1) \\ &= E(I_i) + (\boldsymbol{\Omega}_{<i}^{-1} \cdot \boldsymbol{\Omega}_{i,<i})'(1 - E(I_1),\ 1 - E(I_2) ... 1 - E(I_{i-1}))' \qquad (9) \\ &= \Phi(w_i) + (\boldsymbol{\Omega}_{<i}^{-1} \cdot \boldsymbol{\Omega}_{i,<i})'(1 - \Phi(w_1),\ 1 - \Phi(w_2) ... 1 - \Phi(w_{i-1}))' \end{aligned}$$

Where, $\boldsymbol{\Omega}_{i,<i}$ is a column vector formed from the covariance elements $\mathrm{Cov}(\tilde{I}_i, \tilde{I}_j)$ where j =1, 2….i–1 and $\boldsymbol{\Omega}_{<i}$ is an (i–1) × (i–1) matrix whose (j, k) element is formed by covariance elements $\mathrm{Cov}(\tilde{I}_j, \tilde{I}_k)$ with j and k satisfying the following condition: 1≤ j, k ≤ i–1. The approximation can be interpreted as a linear regression model in which $\tilde{I}_i$ is the dependent variable and the remaining indicator variables, $\tilde{\boldsymbol{I}}_{<i} = (\tilde{I}_1, \tilde{I}_2, ... \tilde{I}_{i-1})$, are the independent variables. Please refer to Bhat (2011a) to get a more detailed treatment on how these matrices are formed. The covariance term can be evaluated as $\mathrm{Cov}(\tilde{I}_j, \tilde{I}_k) = E(\tilde{I}_j \tilde{I}_k) - E(\tilde{I}_j) E(\tilde{I}_k)$. The computation of $E(\tilde{I}_j \tilde{I}_k)$ involves a bivariate CDF whereas that of $E(\tilde{I}_j)$ involves a univariate CDF. So using only univariate and bivariate CDF function evaluations, we have a method for computing normal CDFs of any dimension.

The properties of the CML estimator discussed earlier in terms of the asymptotic normality and the selection criterions are applicable to the MACML estimator also and maybe used without any modification. Since the approximation adopted by Bhat (2011a) relies only on bivariate and univariate standard normal cumulative distribution function computations, the method is computationally efficient. The MACML approach can be applied using simple optimization software for likelihood estimation. It also represents a conceptually simpler alternative to simulation techniques, and has the advantage of reproducibility of the results.

1.4 RESEARCH OBJECTIVE AND DISSERTATION OUTLINE

The use of mixed GEV kernel models for modeling various components of the transportation systems is now a standard practice, thanks to the advances in the MSL estimation method and the suitability of GEV kernel models to such estimations. The analyst's choice between a GEV kernel and a probit kernel was governed by estimation consideration rather than behavioral or model fit consideration. This has resulted in most of the applied work in the area of unordered choice modeling using GEV kernel models. However, simulation methods for estimating GEV kernel models break down when spatial interactions are introduced in the formulation and consequently there is a lack of spatial models to analyze unordered choices. The main goal of this dissertation is, therefore, to bring the probit kernel models back to use through the use of the MACML estimation technique, especially for modeling problems for which spatial interaction effects are critical. This is achieved both by the use of simulation experiments to compare the MACML approach to the MSL method and by the formulation and estimation of spatial models for different empirical problems involving unordered choices.

For my dissertation, I have four objectives. The first is to evaluate the performance of the MACML method to estimate unordered response models by undertaking a Monte Carlo simulation exercise. Estimations are performed for mixed cross-sectional and panel multinomial probit models. MSL estimations are also performed to compare the results of the two methods. The second objective is to formulate and estimate a spatial and temporal unordered discrete choice model and apply this model to a land use change context and to the mode choice decision of school children. While land-use analysis has been an active research area for several decades, the more recent public availability of longitudinal and high resolution spatial land-use data (collected using aerial photography, remote-sensing, and/or real-estate appraisal information) has facilitated the estimation of rich empirical models of land-use. In particular, earlier land-use models used a coarse resolution for the spatial unit of analysis, employing an appropriate transformation of aggregate fractions in each of several types of land-uses as the dependent variables in a linear regression setting. On the other hand,

thanks to new spatial data collection and assembly technologies, more recent land-use models have been able to use a very fine resolution for the spatial unit of analysis, employing discrete indicators for the type of land-use of each spatial unit as the dependent variables in a discrete choice model. Similarly, in children's school mode choice studies it is loosely acknowledged that spatial interaction effect plays a role, however, few studies incorporate spatial interactions systematically in their analysis. Spatial interaction may occur in two possible ways – across spatial units (zones, neighborhoods, tracts, blocks) because units that are closer to one another share some common unobserved attributes, and/or across behavioral units (individuals, households) because behavioral units that are closer to one another in space may share common unobserved attributes that affect the way they behave. The model is formulated to account for such interactions.

The first two objectives discussed above helps towards benchmarking the MACML estimation method and applying the MAMCL method to estimate unordered choice models with spatial interactions. The next objective addresses one of the drawbacks of the original MACML method that they could handle only normal mixing distributions. Non-normal mixing distributions would have necessitated the use of an appropriate finite normal mixture distribution. Thus the third objective is to extend the MACML method to allow for the estimation of non-normal mixing distributions on model parameters. We propose the use of the multivariate skew-normal distribution function which is a generalized form of the normal distribution. Finally, the fourth objective is to propose an improvement to the MACML method by incorporating a second order MVNCD function that is more accurate and evaluate its performance in estimating parameters for a variety of model structures.

The rest of the dissertation is structured as follows. Chapter 2 presents the results of the study undertaken to evaluate the performance of MACML method in estimating various types of models. The third chapter formulates and estimates a school mode choice model that is capable of capturing the unobserved spatial interaction effects that may potentially influence household decision-making processes when choosing a mode of

transportation for children's trips to and from school. The fourth chapter presents an empirical discrete land-use model within a spatially explicit economic structural framework for land-use change decisions. In the fifth chapter we propose the use of the multivariate skew-normal distribution function to accommodate non-normal mixing in MNP models within a MACML estimation framework. The sixth chapter presents the improved MACML estimation approach which incorporates the second order MVNCD approximation. Finally, the seventh chapter summarizes the conclusions of the dissertation and lists some avenues of future research.

CHAPTER 2: A SIMULATION EVALUATION OF THE MAXIMUM APPROXIMATE COMPOSITE MARGINAL LIKELIHOOD ESTIMATOR FOR MIXED MULTINOMIAL PROBIT MODELS

The material in this chapter is drawn substantially from the following published paper.

Bhat, C.R., Sidharthan, R., (2011) A simulation evaluation of the maximum approximate composite marginal likelihood (MACML) estimator for mixed multinomial probit models. *Transportation Research Part B* 45(7), 940-953. [1]

2.1 INTRODUCTION

Consider the following random-coefficients formulation in which the utility that an individual *q* associates with alternative *i* is given by:

$$U_{qi} = \boldsymbol{\beta}_q' \boldsymbol{x}_{qi} + \varepsilon_{qi} \quad (1)$$

where $\boldsymbol{x}_{qi}$ is a (K×1)-column vector of exogenous attributes, and $\boldsymbol{\beta}_q$ is an individual-specific (K×1)-column vector of corresponding coefficients that is a realization from a multivariate normal density function with mean vector $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\Omega}$. ε_{qi} is assumed to be an independently and identically distributed (across alternatives and across individuals) error term, which is also independent of the covariate vector $\boldsymbol{x}_{qi}$. If ε_{qi} is normally distributed with a mean zero and variance of one-half, then the likelihood contribution of individual *q* who chooses alternative *m* is:

$$L_q = \int_{\boldsymbol{\beta}=-\infty}^{\infty} \left\{ \int_{\lambda=-\infty}^{\infty} \left(\prod_{i \neq m} \left[\Phi\left\{ \left[-\sqrt{2}(\boldsymbol{\beta}' \boldsymbol{z}_{qim}) \right] + \lambda \right\} \right] \right) \phi(\lambda) d\lambda \right\} f(\boldsymbol{\beta} \mid \boldsymbol{b}, \boldsymbol{\Omega}) d\boldsymbol{\beta}, \quad (2)$$

[1] The author of this dissertation collaborated with the coauthor on the methodological and technical aspects of the paper.

where $\boldsymbol{z_{qim} = x_{qi} - x_{qm}}$, $\Phi(\cdot)$ is the univariate cumulative distribution function and $\phi(\cdot)$ is the univariate normal density function. In the case of panel data, the utility structure may be written with the inclusion of choice occasion *t* as:

$$U_{qit} = \boldsymbol{\beta'_q x_{qit}} + \varepsilon_{qit}. \tag{3}$$

In this case, the individual likelihood contribution of individual *q* choosing alternative m_t at choice occasion *t* when ε_{qit} is normally distributed, is:

$$L_q = \int_{\boldsymbol{\beta}=-\infty}^{\infty} \prod_{t=1}^{T} \left[\int_{\lambda=-\infty}^{\infty} \left(\prod_{i \neq m_t} \left[\Phi\left\{ \left[-\sqrt{2}(\boldsymbol{\beta' z_{qim_t t}}) \right] + \lambda \right\} \right] \right) \phi(\lambda) d\lambda \right] f(\boldsymbol{\beta} \mid \boldsymbol{b}, \boldsymbol{\Omega}) d\boldsymbol{\beta}, \tag{4}$$

where $\boldsymbol{z_{qim_t t} = x_{qit} - x_{qmt}}$.

Finally, in the case of panel data, and when the random coefficients have both an intra-individual and inter-individual random component (see Bhat and Castelar, 2002; Bhat and Sardesai, 2006; Hess and Rose, 2009), the utility structure may be written as:

$$U_{qit} = \boldsymbol{\beta'_{qt} x_{qit}} + \varepsilon_{qit}, \tag{5}$$

where $\boldsymbol{\beta_{qt} = \beta_q + \tilde{\beta}_{qt}}$, $\boldsymbol{\beta_q} \sim N(\boldsymbol{b}, \boldsymbol{\Omega})$, $\tilde{\boldsymbol{\beta}}_{qt} \sim N(0, \tilde{\boldsymbol{\Omega}})$.

In this case, when ε_{qit} is normally distributed,

$$L_q = \int_{\boldsymbol{\beta}=-\infty}^{\infty} \left[\prod_{t=1}^{T} \left(\int_{\tilde{\boldsymbol{\beta}}=-\infty}^{\infty} \left\{ \int_{\lambda=-\infty}^{\infty} \left(\prod_{i \neq m_t} \left[\Phi\left\{ \left[-\sqrt{2}(\boldsymbol{\beta'} + \tilde{\boldsymbol{\beta}}') \mathbf{z}_{\mathbf{qim_t t}} \right] + \lambda \right\} \right] \right) \phi(\lambda) d\lambda \right\} f(\tilde{\boldsymbol{\beta}} \mid \tilde{\boldsymbol{\Omega}}) d\tilde{\boldsymbol{\beta}} \right) \right] \tag{6}$$
$$f(\boldsymbol{\beta} \mid \boldsymbol{b}, \boldsymbol{\Omega}) d\boldsymbol{\beta}.$$

The likelihood contribution of individual *q* in Equations (2), (4), and (6) entails the evaluation of an analytically intractable function with multidimensional integrals. This has led to the development of various simulation techniques in high dimensions as part of a maximum simulated likelihood (MSL) estimation approach. Unfortunately, for many practical situations, the computational cost to ensure good asymptotic MSL estimator properties can be prohibitive and literally infeasible (in the context of the computation resources available and the time available for estimation) as the number of dimensions of integration increases.

Bhat (2011a) proposed the use of an alternative maximum approximate composite marginal likelihood (MACML) estimator within the class of frequentist estimators for the estimation of multinomial probit (MNP) models. Bhat's MACML estimator is based solely on univariate and bivariate cumulative normal distribution evaluations, regardless of the dimensionality of integration. This should substantially reduce computation time compared to more cumbersome simulation techniques to evaluate multidimensional integrals. At the same time, the MACML estimator retains the properties of being consistent and asymptotically normally distributed.

The specific objectives of this study are motivated by the discussion above. The first objective is to examine the ability of the MACML estimator to recover parameters from finite samples in mixed cross-sectional and panel multinomial probit models. Simulated data sets with known underlying model parameters are used to evaluate the MACML approach. The second, related, objective is to compare the performance of the MACML approach with the MSL approach in mixed MNP simulations when the MSL approach is feasible. In doing so, we examine the relative ability of the two approaches to recover parameters and the computation time of the two approaches.

The rest of the chapter is structured as follows. Section 2.2 presents the experimental design for the simulation experiments and Section 2.3 presents the results. Section 2.4 concludes the chapter by highlighting important findings.

2.2 EXPERIMENTAL DESIGN

In the simulation set-up to examine the performance of the MSL and MACML inference approaches, we consider the case of five alternatives with five independent variables. For all the datasets generated in the experimental design, the values of each of the five independent variables for the alternatives are drawn from a standard univariate normal distribution. For the cross-sectional data set, we generate a sample of 5000 realizations of the five independent variables corresponding to 5000 individuals, while, for the panel data set, we generate a sample of 2500 realizations of the five independent variables corresponding to a situation where 500 individuals each have five choice occasions for a

total of 2500 choice occasions. We allow random coefficients on all the five independent variables. This leads to a five-dimensional integral in the mixed model. In the subsequent three sections, we discuss the set-up for each of the following three cases in more detail: (1) cross-sectional random coefficients, (2) panel inter-individual coefficients and (3) panel intra-individual and inter-individual random coefficients.

2.2.1 Cross-Sectional Random Coefficients Model Structure

In the cross-sectional case, the coefficient vector $\boldsymbol{\beta}_q$ for individual q is assumed to be a realization from a multivariate normal distribution with a mean vector $\boldsymbol{b}$ = (1.5, –1, 2, 1, –2) and covariance matrix $\boldsymbol{\Omega}$. Two specifications for $\boldsymbol{\Omega}$ are considered. The first specification, which we label as the diagonal covariance specification, assumes independence among the random coefficients; that is, the matrix $\boldsymbol{\Omega}$ is assumed to be diagonal. This specification has been frequently used in the literature. The entries along the diagonal are set to the value of 1 in the experimental design. This first specification entails the estimation of five parameters in the covariance matrix. The second specification, which we label as the non-diagonal covariance specification, allows the random coefficients to be correlated. In this specification, we specify the matrix $\boldsymbol{\Omega}$ to be as follows:

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & -0.50 & 0.25 & 0.75 & 0 \\ -0.50 & 1 & 0.25 & -0.50 & 0 \\ 0.25 & 0.25 & 1 & 0.33 & 0 \\ 0.75 & -0.50 & 0.33 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This positive definite non-diagonal specification involves the estimation of 10 covariance matrix parameters. Finally, values for the error terms ε_{qi} (q = 1, 2, …, Q; i = 1, 2, …, I) in Equation (1) are generated from a univariate normal distribution with a variance of 0.5, leading to the mixed MNP model structure. The alternative with the highest utility for each observation is then identified as the chosen alternative. The above data generation process is undertaken 20 times with different realizations of the $\boldsymbol{\beta}_q$ vector and the error

term ε_{qi} to generate 20 different data sets each for the diagonal specification and the non-diagonal specification of the $\boldsymbol{\Omega}$ matrix.

The MSL and MACML estimators are applied to each data set to estimate data specific values of $\boldsymbol{b}$ and $\boldsymbol{L}$ ($\boldsymbol{\Omega} = \boldsymbol{LL}'$, where $\boldsymbol{L}$ is the lower Cholesky decomposition of $\boldsymbol{\Omega}$; note that it is the Cholesky parameters that are estimated to ensure the positive definiteness of the variance-covariance matrix $\boldsymbol{\Omega}$). In the case of the diagonal covariance specification, $\boldsymbol{L}$ is also a diagonal matrix with entries of '1' along the diagonal. The MSL estimator is applied to each dataset 10 times with different (independent) draws for the random coefficients for each individual. This allows us to estimate the simulation error in the MSL case by computing the standard deviation of estimated parameters among the 10 different estimates on the same data set. Similarly, for the MACML approach, the approximation error is obtained by computing the standard deviation of estimated parameters among the 10 different estimates on the same data set by using different permutations to decompose the multivariate normal cumulative distribution (MVNCD) function into a product sequence of marginal and conditional probabilities (see Section 2.1 of Bhat, 2011a).

For the MSL estimation, we use draws from the Halton sequence for the random coefficients vector $\boldsymbol{\beta}_q$, because it is the most commonly used QMC sequence in the literature. While some other QMC systems have been shown to provide better results for a given number of draws, the Halton has the advantage of very easy generation. Thus, as indicated by Sandor and Train (2004), one can generate many more draws per individual of the Halton sequence than other QMC sequences for the same amount of time. Within the context of Halton draws, we experimented with different kinds of scramblings and randomizations of the Halton sequence (see Bhat, 2003 and Sivakumar *et al.*, 2005 for a review of these scrambling and randomization techniques). The experiments indicated that the best performance was obtained using a procedure that combined Bratten-Weller scrambling with the Tuffin randomization, further enhanced by the random assignment of Halton dimensions to coefficients. Also, while a higher number of draws per individual

(based on the combination scrambling/randomization discussed above) generally provided improved results, we used 250 draws per individual, which is more than what is typically used in most applications of the MSL procedure. Further, with a total of 400 total estimations for the cross-sectional random coefficients case (20 simulation runs for each of 10 different data samples for each of the diagonal and non-diagonal covariance case), an important factor was to keep the computation cost per estimation to a reasonable amount of time (even with 250 draws per individual, the total computer time for the 400 estimations was over 800 hours, as we discuss in more detail later). Finally, note that one has to integrate out the inner one-dimensional integral over the scalar λ that is distributed standard normal (see Equation (2)). While this integration can also be performed using QMC draws, we undertake this inner one-dimensional integration using the more efficient hermite quadrature technique with 10 quadrature points.

For the MACML method, a single random permutation is generated for each individual (the random permutation varies across individuals, but is the same across iterations for a given individual), and the multivariate normal cumulative distribution (MVNCD) function is approximated using the resulting conditional probability sequence. We used different numbers of random permutations per individual to approximate the MVNCD function corresponding to the individual likelihood contribution. However, there was hardly any difference between using a single permutation and higher numbers of permutations, and hence we used a single permutation per individual (in one of the 400 estimations undertaken in the cross-sectional case, using two permutations per individual instead of a single permutation provided stability to the iterations).

2.2.2 Panel Inter-Individual Random Coefficients

As in the cross-sectional case, for the panel case too, we consider both a diagonal specification as well as a non-diagonal specification for the $\boldsymbol{\beta}_q$ random coefficient vector, with the mean vector and the covariance matrix of $\boldsymbol{\beta}_q$ identical to the cross-sectional case. The difference is that we generate only 500 vectors of coefficients, one vector for each of the 500 individuals. The same individual-specific coefficient vector is

applied to all 5 choice occasions of the individual. The values for the error terms ε_{qit} are generated from a univariate normal distribution with a variance of 0.5, and the alternative with the highest utility is designated as the chosen alternative at each choice occasion for each individual.

The data generation process is undertaken 10 times with different sets of 500 realizations of the $\boldsymbol{\beta_q}$ vector and 2500 realizations of the error term ε_{qit} to obtain 10 different data sets (we used fewer data samples and fewer total observations for the panel case compared to the cross-sectional case because of the increased computational costs for panel data relative to cross-sectional data). The MSL and MACML estimation procedures are applied to each data set. For the MSL approach, we decided to ignore simulation error and estimated only a single set of parameters for each data set using 250 Halton draws because of the computation time involved. Also, we observed during the MSL runs that the analytic gradient function was not returning accurate values consistent with the likelihood function for 10 quadrature points when integrating out λ in Equation (4). This is not surprising, since the product across choice occasions of the same individual is now within the integration for λ. The net result was that the convergence process would get stuck because of the inaccuracy. So, we had to increase the accuracy of the gradient procedure by increasing the number of hermite quadrature points to 40 in the panel case. For the MACML case, we estimated the approximation error by estimating the model 10 times for each data set with different sets of permutations (as in the cross-sectional case). We tested the performance of the MACML method by using both a single permutation per individual as well as two permutations per individual, and found (as in the cross-sectional case) that the performance improvement was rather marginal.

2.2.3 Panel Intra-Individual and Inter-Individual Random Coefficients

This estimation involves the generation of the 500 vector realizations of coefficients for $\boldsymbol{\beta_q}$ as earlier from the multivariate normal distribution with a mean vector $\boldsymbol{b}$ and covariance $\boldsymbol{\Omega}$. In addition, 2,500 vectors of coefficients for $\tilde{\boldsymbol{\beta}}_{\boldsymbol{qt}}$ (see Equation (5)) are generated from the multivariate normal distribution with a mean vector of 0 and

covariance $\tilde{\boldsymbol{\Omega}}$. As in the earlier cases, we considered both a diagonal specification for $\boldsymbol{\Omega}$ and $\tilde{\boldsymbol{\Omega}}$, as well as non-diagonal specifications for both covariance matrices. The diagonal specification involved draws for $\boldsymbol{\beta}_q$ and $\tilde{\boldsymbol{\beta}}_{qt}$ from standard and independently normally distributions, while the non-diagonal covariance specification for $\boldsymbol{\beta}_q$ was the same as in Section 2.2.2 and the non-diagonal covariance specification for $\tilde{\boldsymbol{\beta}}_{qt}$ was as below:

$$\tilde{\boldsymbol{\Omega}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0.5 & 0.5 & 0.5 \\ 0 & 0.5 & 1 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 0.5 & 1 \end{bmatrix}$$

The sum of $\boldsymbol{\beta}_q$ and $\boldsymbol{\beta}_{qt}$ realizations are then applied to the independent variable vector for each individual's choice occasion to identify the alternative with highest utility. Everything else remains identical to Section 2.2.2.

2.3 PERFORMANCE COMPARISON BETWEEN THE MSL AND MACML APPROACHES

In this section, we first identify a number of performance measures and discuss how these are computed for the MSL approach and the MACML approach. The subsequent sections present the simulation and computational results.

2.3.1 Performance Measures

The steps discussed below for computing performance measures are for a specific correlation matrix pattern. We discuss the approach first for the cross-sectional random coefficients case, and then indicate the minor modifications for the two panel random coefficients cases.

MSL Approach

(1) Estimate the MSL parameter estimates for each data set *s* (*s* = 1, 2, …, 20) and for each of the 10 independent scrambled and randomized Halton draws, and obtain the time to obtain the convergent values and the standard errors. Obtain the mean time for convergence (TMSL) and standard deviation of convergence time across the 200

runs for each correlation pattern The 200 runs correspond to 10 runs for each of 20 data sets. The time to convergence includes the time to compute the covariance matrix of parameters and the corresponding parameter standard errors. All estimations are started with the true parameter values as the starting values. While multiple computers had to be used for the many different runs undertaken in this study, all the run times were carefully scaled to the equivalent time on a desktop computer with 3GHz Quad core processor and 8GB of RAM. The scaling was based on extensive experimentation on different computers.

(2) For each data set *s* and draw combination, estimate the standard errors (s.e.) of parameters (using the sandwich estimator; see McFadden and Train, 2000).

(3) For each data set *s*, compute the mean estimate for each model parameter across the draws. Label this as MED, and then take the mean of the MED values across the data sets to obtain **a mean estimate**. Compute the **absolute percentage bias** (APB) as: $APB = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100$.[2]

(4) For each data set *s*, compute the median s.e. for each model parameter across the 10 draws. Call this MSED, and then take the mean of the MSED values across the 20 data sets and label this as **the asymptotic standard error** (essentially this is the standard error of the distribution of the estimator as the sample size gets large). Note that we compute the median s.e. for each model parameter across the draws and label it as MSED rather than computing the mean s.e. for each model parameter across the draws. This is because, for some draws, the estimated standard errors turned out to be rather large relative to other independent standard error estimates for the same dataset. Note that the mean asymptotic standard error is a theoretical approximation to the finite sample standard error.

(5) Next, for each data set *s*, compute the simulation standard deviation for each parameter as the standard deviation in the estimated values across the independent

[2] In case a true parameter value is zero, the APB is computed by taking the difference of the mean estimate from the true value (= 0), dividing this difference by the value of 1 in the denominator, and multiplying by 100.

draws (about the MED value). Call this standard deviation as SIMMED. For each parameter, take the mean of SIMMED across the different data sets. Label this as the **simulation standard error** for each parameter.

(6) For each parameter, compute a **simulation adjusted asymptotic standard error** as follows: $\sqrt{(\text{asymptotic standard error})^2 + (\text{simulation standard error})^2}$

MACML Approach

(1) Estimate the MACML parameters for each data set *s* and for each of 10 independent sets of permutations for computing the approximation for the likelihood function contribution of each individual. Obtain the time to get the convergent values (including the time to obtain the covariance matrix based on the inverse of the Godambe information matrix and the corresponding standard errors). Determine the mean time for convergence (TMACML) across the 200 estimation runs for each correlation pattern. As in the MSL runs, estimations were begun with the true values as the starting values, and the run times on different computers were scaled to an equivalent time on the baseline computer.

(2) For each data set *s*, estimate the standard errors (s.e.) (using the Godambe estimator; see Bhat, 2011a).

(3) For each data set *s*, compute the mean estimate for each model parameter across the 10 random permutations used. Label this as MED, and then take the mean of the MED values across the data sets to obtain **a mean estimate**. Compute the **absolute percentage bias** (APB) as in the MSL case.

(4) For each data set *s*, compute the median s.e. for each model parameter across the 10 draws. Call this MSED, and then take the mean of the MSED values across the 20 data sets and label this as **the asymptotic standard error.**

(5) Next, for each data set *s*, compute the approximation standard deviation for each parameter as the standard deviation in the estimated values across the independent permutations (about the MED value). Call this standard deviation as APPMED. For each parameter, take the mean of APPMED across the different data sets. Label this as the **approximation standard error** for each parameter.

(6) For each parameter, compute an **approximation adjusted asymptotic standard error** as follows: $\sqrt{(\text{asymptotic standard error})^2 + (\text{approximation standard error})^2}$.

The procedure above is applied for the cross-sectional random-coefficients case. For the panel inter-individual random coefficients case, and the panel inter-individual and intra-individual random coefficients case, the same approach as above is used, except that we generate only 10 datasets instead of 20 datasets. Also, only one MSL run is undertaken for each dataset, and so no simulation standard errors are computed for the MSL. This is because of the computational cost involved for each MSL run in the panel cases. However, we do compute the approximation standard errors for the MACML estimations by running 10 independent sets of permutations for each of the 10 datasets.

2.3.2 Results

2.3.2.1 The Cross-Sectional Random Coefficients (CSRC) Model

Table 2.1a presents the results for the CSRC model with a diagonal covariance matrix, and Table 2.1b presents the corresponding results for the CSRC model with a non-diagonal covariance matrix.

The Diagonal Case

The results in Table 2.1a for the diagonal case indicate that both the MSL and the MACML method do reasonably well in recovering the parameters, as can be observed by comparing the mean estimate of the parameters with the true values (see the column titled "parameter estimates). The absolute percentage bias (APB) ranges from 7.3% to 13.3% (overall mean value of 9.8% across parameters - see the row of the table labeled "Overall mean value across parameters" and the column titled "absolute percentage bias") for the MSL approach, and from 0.2% to 5.9% (overall mean value of 2.5% across parameters) for the MACML approach. Clearly, the MACML is able to recover parameters much more accurately than the MSL approach. For both the MSL and MACML methods, the APB values are generally somewhat smaller for the mean values of the distributions of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the *b* values in the table) than for the standard deviations of the distribution of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the σ parameters in the table). Also,

there is more variation in the APB values among the σ parameters than among the b values, suggesting that the log-likelihood function is relatively flat for different values of standard deviations, leading to somewhat more difficulty in accurately recovering the standard deviation parameters. The sampling standard error values of the parameters indicate good efficiency of both the MSL and MACML estimators, with the asymptotic standard error being only about a tenth of the mean values of the estimator. The asymptotic standard error values may appear to suggest that the MSL estimator is marginally more efficient than the MACML estimator, given that the asymptotic standard errors from the MSL are slightly lower than from the MACML approach (the mean asymptotic standard error from the MSL method is 0.121, while the mean asymptotic standard error from the MACML method is 0.151). However, note that the lower standard errors from the MSL method are simply an artifact of the underestimation in recovering the true values of the parameters, which translates to consistently lower values of the mean parameter estimates from the MSL approach relative to the MACML method. In fact, in the MSL runs where the estimated parameters were of the order of the mean estimates from the MACML method, the corresponding MSL asymptotic standard errors were of the same order of magnitude as from the MACML method. Finally, the reader will note that the simulation standard error estimates are smaller than the sampling standard errors in the MSL approach, and similarly the approximation standard error estimates are smaller than the sampling standard errors in the MACML approach. On average, the simulation standard error is about 37% of the sampling standard error in the MSL case, while the approximation standard error is only about 13% of the sampling standard error in the MACML case. It is indeed quite remarkable that the approximation standard error with just a single permutation for approximating the likelihood function contribution of each individual in the MACML approach should be lower than the simulation standard error with 250 Halton draws per individual in the MSL approach. The final column provides the simulation-adjusted asymptotic standard error for the MSL case and the approximation-adjusted asymptotic standard error for the MACML case.

Table 2.1a Evaluation of the ability to recover true parameters for the cross-sectional diagonal case

Parameter	True Value	MSL Method					MACML Method				
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute % Bias	Asymptotic SE	Simul. SE	Simul. Adj. Asymptotic SE	Mean Estimate	Absolute % Bias	Asymptotic SE	Approx. SE	Approx. Adj. Asymptotic SE
Mean values of the β vector											
$b1$	1.500	1.366	9.0%	0.129	0.050	0.139	1.472	1.9%	0.167	0.022	0.169
$b2$	-1.000	-0.906	9.4%	0.089	0.033	0.095	-0.976	2.4%	0.113	0.014	0.114
$b3$	2.000	1.801	10.0%	0.167	0.066	0.180	1.940	3.0%	0.218	0.028	0.219
$b4$	1.000	0.906	9.4%	0.089	0.034	0.095	0.977	2.3%	0.114	0.014	0.114
$b5$	-2.000	-1.820	9.0%	0.170	0.067	0.182	-1.960	2.0%	0.220	0.028	0.222
Standard deviations of the β vector											
$\sigma1$	1.000	0.885	11.5%	0.111	0.038	0.117	0.958	4.2%	0.135	0.017	0.137
$\sigma2$	1.000	0.906	9.4%	0.111	0.040	0.118	0.984	1.6%	0.136	0.016	0.137
$\sigma3$	1.000	0.867	13.3%	0.112	0.041	0.119	0.941	5.9%	0.135	0.017	0.136
$\sigma4$	1.000	0.904	9.6%	0.111	0.040	0.118	0.982	1.8%	0.136	0.017	0.137
$\sigma5$	1.000	0.927	7.3%	0.117	0.041	0.124	1.002	0.2%	0.140	0.016	0.141
Overall Mean Value		-	9.8%	0.121	0.045	0.129	-	2.5%	0.151	0.019	0.153
Mean Time		66.09					1.96				
Std. dev of Time		10.87					0.42				
% of Runs Converged		100%					100%				

Table 2.1b Evaluation of the ability to recover true parameters for the cross-sectional non-diagonal case

Parameter	True Value	MSL Method					MACML Method				
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute % Bias	Asymptotic SE	Simul. SE	Simul. Adj. Asymptotic SE	Mean Estimate	Absolute % Bias	Asymptotic SE	Approx. SE	Approx. Adj. Asymptotic SE
Mean values of the β vector											
$b1$	1.500	1.374	8.4%	0.133	0.049	0.142	1.443	3.8%	0.147	0.022	0.148
$b2$	-1.000	-0.912	8.8%	0.093	0.037	0.100	-0.959	4.1%	0.102	0.014	0.103
$b3$	2.000	1.830	8.5%	0.174	0.068	0.187	1.923	3.8%	0.191	0.029	0.193
$b4$	1.000	0.914	8.6%	0.092	0.032	0.097	0.958	4.2%	0.101	0.014	0.102
$b5$	-2.000	-1.849	7.6%	0.176	0.068	0.189	-1.941	3.0%	0.194	0.028	0.196
Cholesky parameters characterizing the covariance matrix of the β vector											
$l11$	1.000	0.909	9.1%	0.112	0.040	0.119	0.959	4.1%	0.119	0.017	0.120
$l12$	-0.500	-0.463	7.3%	0.085	0.029	0.090	-0.472	5.6%	0.085	0.010	0.085
$l13$	0.250	0.231	7.5%	0.089	0.036	0.096	0.233	6.7%	0.087	0.009	0.088
$l14$	0.750	0.689	8.2%	0.092	0.028	0.097	0.707	5.7%	0.095	0.013	0.096
$l15$	0.000	0.006	0.6%	0.086	0.040	0.095	0.015	1.5%	0.088	0.008	0.089
$l22$	0.866	0.756	12.7%	0.109	0.043	0.117	0.809	6.5%	0.116	0.017	0.117
$l23$	0.433	0.431	0.5%	0.105	0.050	0.117	0.436	0.6%	0.100	0.012	0.101
$l24$	-0.144	-0.149	3.6%	0.101	0.041	0.109	-0.170	17.8%	0.093	0.010	0.094
$l25$	0.000	-0.021	2.1%	0.101	0.055	0.115	-0.019	1.9%	0.098	0.010	0.099
$l33$	0.866	0.750	13.4%	0.130	0.073	0.149	0.812	6.3%	0.131	0.019	0.132
$l34$	0.237	0.242	2.0%	0.112	0.055	0.125	0.259	9.3%	0.106	0.011	0.106
$l35$	0.000	-0.031	3.1%	0.120	0.081	0.145	-0.029	2.9%	0.116	0.011	0.117
$l44$	0.601	0.464	22.9%	0.126	0.085	0.152	0.531	11.6%	0.125	0.015	0.126
$l45$	0.000	-0.053	5.3%	0.168	0.134	0.214	-0.053	5.3%	0.171	0.017	0.172
$l55$	1.000	0.885	11.5%	0.125	0.089	0.153	0.956	4.4%	0.136	0.018	0.137
Overall Mean Value		-	7.6%	0.116	0.057	0.130	-	5.5%	0.120	0.015	0.121
Mean Time		174.32					5.19				
Std. dev of Time		28.13					0.84				
% Runs Converged		100%					100%				

These values are very close to the unadjusted asymptotic standard error in the MSL case and the unadjusted asymptotic standard error in the MACML case, once again indicating that the simulation and approximation errors are small relative to the sampling errors.

The time to convergence for the MSL estimation has a mean value of 66.1 minutes with a standard deviation of about 11 minutes. On the other hand, the time to convergence for the MACML estimation has a mean value of 1.96 minutes with a standard deviation of about 0.5 minutes. This indicates that the MACML method is about 33 times faster than the MSL estimation. Further, note that the MACML method is actually much more effective than suggested by this factor of 33, because it produces more accurate estimates than the MSL estimates. Some further explorations indicated that, even if the analyst increased the number of Halton draws to 450 per individual, the resulting APB (computed from 10 runs on 10 datasets of the 20 datasets) is in the order of 6.8% (relative to 2.5% for the MACML method), and the mean amount of time for convergence with 450 Halton draws is about 107 minutes, suggesting a time efficiency factor of well over 50 for the MACML method relative to the MSL method. This is indeed a phenomenal computational efficiency jump. As the number of random coefficients increase beyond five, one can only expect a further increase in the computational time advantage of the MACML over the MSL estimation approach.

The Non-Diagonal Case

The results in Table 2.1b provide information on the true mean values of the distribution of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the *b* values in the table) and the Cholesky-decomposed parameters characterizing the covariance matrix of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the *l* values in the table). The table also provides information on the mean estimates and the standard error estimates of the above parameters from the MSL and MACML approaches.

As in the diagonal case, the MSL and MACML methods perform well in terms of recovering the true parameter values. In fact, the MSL does marginally better than in the diagonal case, with the absolute percentage bias (APB) ranging from 0.5% to 22.9%,

with an overall mean APB value of 7.6%. However, the MACML model still outperforms the MSL method, with an APB ranging from 0.6% to 17.8% and a mean APB value of 5.5%. As in the diagonal case, there is more stability in the APB values across the mean values of the distribution of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the *b* values in the table) than for the Cholesky parameters characterizing the covariance matrix of the distribution of the $\boldsymbol{\beta}$ parameter vector (*i.e.*, the *l* parameters in the table). The asymptotic standard error estimates again indicate good efficiency of both the MSL and MACML estimators, with the asymptotic standard error being only about a tenth of the mean values of the estimator for the *b* values. Of course, the asymptotic standard errors for the *l* parameters are a higher fraction of the mean estimates for these parameters, which is to be expected since many more parameters are being estimated in the covariance matrix. Between the MSL and the MACML estimators, the asymptotic standard errors are very similar in this non-diagonal case, with the mean standard error being 0.116 in the MSL case and 0.120 in the MACML case. This is because the MSL provides estimates that are closer to the true values, and to the values from the MACML estimation, unlike in the diagonal case. In terms of the simulation standard error in the MSL case and the approximation standard error in the MACML case, these are once again only a fraction of the sampling errors. However, as in the diagonal case, the simulation standard errors for the MSL case are much higher than the approximation standard errors from the MACML case. In particular, the simulation standard error is, on average, 49% of the sampling standard error in the MSL case, while the approximation error is, on average, only 12.5% of the sampling standard error in the MACML case.

The time to convergence for the MSL estimation has a mean value of 174.3 minutes (almost three hours) with a standard deviation of 28 minutes. In contrast, the time to convergence for the MACML estimation has a mean value of 5.20 minutes with a standard deviation of about 0.9 minutes. These results indicate that the MACML method is, once again and coincidentally, about 33 times faster than the MSL estimation with 250 Halton draws. However, for an apples-to-apples comparison, one needs to improve the estimation with MSL, which we attempted to do by increasing the number of Halton

draws. When using 450 draws per individual to estimate parameters using 10 runs on 10 datasets of the 20 datasets, the mean APB value turned out to be 9.6% with a mean time of convergence of about 380 minutes. This mean APB is higher compared to the 250 draws per individual case, and raises a yet unexplored issue with the Halton and related QMC draws. That is, the effectiveness of the standard and/or scrambled QMC draws may not be stable as the number of draws is increased, because the cycling of the QMC sequences may lead to poor coverage of the multivariate space for any given individual for specific numbers of draws.

2.3.2.2 Panel Inter-Individual Random Coefficients (PIRC) model

The panel estimations are undertaken with about half the number of total observations as the cross-sectional estimations, so that the computation time can be kept to a reasonable time with the MSL approach. Further, the covariance matrix for the PIRC model is estimated based on individual-level random heterogeneity, based on 500 distinct draws (one for each of the 500 individuals). Also, to keep the computation time reasonable, we use only 10 datasets in the panel case. As a result, one may expect the level of performance of the MACML and the MSL to be, in general, somewhat lower than the cross-sectional case. In the next two sections, we discuss the results for the diagonal and non-diagonal cases.

The Diagonal Case

The results in Table 2.2a for the diagonal case indicate that, for the MSL estimation approach, the absolute percentage bias (APB) ranges from 16.0% to 19.9%, with an overall mean APB of 17.1%. The corresponding APB values for the MACML approach range from 5.8% to 12.4%, with an overall mean value of 8.0%. As in the cross-sectional case, the MSL estimation is undertaken with 250 scrambled and randomized Halton draws, while the MACML estimation is undertaken with a single randomized permutation (except in 6 of the 100 cases, where two randomized permutations provided stability).[3] Clearly, the MACML is able to recover parameters more accurately than the

[3] All time computations discussed later for the one-permutation case include the times for these 6 cases that used two permutations per individual.

MSL approach. The seemingly lower asymptotic standard errors of the MSL approach is again an artifact of the substantial underestimation of parameter values in the MSL approach. The approximation standard error estimates are smaller than the sampling standard errors in the MACML approach. On average, the approximation standard error is only about 28% of the sampling standard error in the MACML case.[4]

The time to convergence for the MSL estimation has a mean value of 96.3 minutes with a standard deviation of about 11 minutes. On the other hand, the time to convergence for the MACML estimation has a mean value of 12.4 minutes with a standard deviation of about 3 minutes. Compared to the cross-section case, the computational efficiency of the MACML over the MSL is not as substantial in the panel diagonal case. This is because of two reasons. The first is that the number of multivariate integrations per likelihood function or gradient iteration is only 500 in the panel case (corresponding to the 500 individuals), and this benefits the MSL approach. Second, in the MACML estimation of the panel case, we consider all the ten pairings of the 5 choice occasions per individual, which increases the number of multi-dimensional integrals to be evaluated using the approximation method to 5000 (500 individuals times 10 pairings per individual). However, the MACML approach still retains a significant computation edge, being about 8 times faster than the MSL approach with 250 randomized and scrambled Halton draws. At the same time, the MACML approach is able to recover parameters much more accurately than the MSL approach. In fact, even when the number of Halton draws was increased to 450 per individual, the MSL had a mean APB of 14.3%, and the corresponding mean time of convergence was 185.6 minutes. This indicates that the actual computation edge of the MACML over the MSL is more than 15-fold.

[4] When we estimated the panel diagonal case with two permutations per individual (rather than one permutation per individual), the mean APB improved marginally to 7.0%. The approximation standard error, on average, turned out to be 15% of the asymptotic sampling standard error in this case. Overall, the results show the ability to recover parameters with small approximation error with just one permutation per individual.

Table 2.2a Evaluation of the ability to recover true parameters for the panel inter-individual random coefficients diagonal case

Parameter	True Value	MSL Method			MACML Method				
		Parameter Estimates		Standard Error Estimates	Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute % Bias	Asymptotic SE	Mean Estimate	Absolute % Bias	Asymptotic SE	Approx. SE	Approx. Adj. Asymptotic SE
Mean values of the β vector									
$b1$	1.500	1.247	16.9%	0.094	1.400	6.6%	0.143	0.043	0.149
$b2$	-1.000	-0.818	18.2%	0.070	-0.914	8.6%	0.102	0.028	0.106
$b3$	2.000	1.660	17.0%	0.111	1.869	6.6%	0.185	0.056	0.194
$b4$	1.000	0.840	16.0%	0.075	0.935	6.5%	0.106	0.029	0.110
$b5$	-2.000	-1.670	16.5%	0.111	-1.870	6.5%	0.184	0.056	0.193
Standard deviations of the β vector									
$\sigma1$	1.000	0.834	16.6%	0.086	0.942	5.8%	0.120	0.034	0.124
$\sigma2$	1.000	0.801	19.9%	0.077	0.876	12.4%	0.111	0.031	0.116
$\sigma3$	1.000	0.844	15.6%	0.093	0.910	9.0%	0.121	0.032	0.125
$\sigma4$	1.000	0.821	17.9%	0.084	0.921	7.9%	0.119	0.031	0.123
$\sigma5$	1.000	0.836	16.4%	0.083	0.900	10.0%	0.119	0.033	0.124
Overall Mean Value		-	17.1%	0.088	-	8.0%	0.131	0.037	0.136
Mean Time		96.26			12.35				
Std. dev of Time		11.13			3.01				
% of Runs Converged		90%			100%				

Table 2.2b Evaluation of the ability to recover true parameters for the panel inter-individual random coefficients non-diagonal case

Parameter	True Value	MSL Method			MACML Method				
		Parameter Estimates		Std. Error Estimates	Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute % Bias	Asymptotic SE	Mean Estimate	Absolute % Bias	Asymptotic SE	Approx. SE	Approx. Adj. Asymptotic SE
Mean values of the β vector									
$b1$	1.500	1.296	13.6%	0.103	1.394	7.1%	0.133	0.034	0.138
$b2$	-1.000	-0.866	13.4%	0.076	-0.943	5.7%	0.099	0.025	0.102
$b3$	2.000	1.747	12.7%	0.133	1.879	6.1%	0.171	0.046	0.177
$b4$	1.000	0.850	15.0%	0.078	0.920	8.0%	0.098	0.023	0.100
$b5$	-2.000	-1.748	12.6%	0.131	-1.879	6.0%	0.170	0.047	0.176
Cholesky parameters characterizing the covariance matrix of the β vector									
$l11$	1.000	0.837	16.3%	0.078	0.914	8.6%	0.105	0.026	0.109
$l12$	-0.500	-0.398	20.4%	0.068	-0.440	12.1%	0.089	0.014	0.090
$l13$	0.250	0.275	9.9%	0.090	0.269	7.5%	0.101	0.012	0.102
$l14$	0.750	0.657	12.5%	0.073	0.690	8.0%	0.094	0.017	0.095
$l15$	0.000	0.011	1.1%	0.073	-0.009	0.9%	0.103	0.017	0.104
$l22$	0.866	0.704	18.7%	0.080	0.745	14.0%	0.095	0.023	0.098
$l23$	0.433	0.314	27.4%	0.095	0.366	15.6%	0.106	0.014	0.107
$l24$	-0.144	-0.075	47.8%	0.065	-0.098	32.2%	0.086	0.014	0.088
$l25$	0.000	0.011	1.1%	0.080	0.023	2.3%	0.103	0.018	0.104
$l33$	0.866	0.764	11.8%	0.106	0.775	10.5%	0.120	0.028	0.124
$l34$	0.237	0.163	31.1%	0.076	0.164	30.7%	0.097	0.017	0.098
$l35$	0.000	-0.015	1.5%	0.093	-0.047	4.7%	0.133	0.021	0.135
$l44$	0.601	0.286	52.4%	0.095	0.498	17.1%	0.110	0.027	0.113
$l45$	0.000	0.184	18.4%	0.104	0.026	2.6%	0.176	0.043	0.182
$l55$	1.000	0.824	17.6%	0.091	0.871	12.9%	0.139	0.042	0.145
Overall Mean Value		-	17.8%	0.090	-	10.6%	0.116	0.025	0.119
Mean Time		192.65			24.41				
Std. dev of Time		52.31			7.81				
% of Runs Converged		50%			100%				

Note also that as soon as slightly more complicated (and more realistic) structures such as autoregressive random coefficients over choice occasions, or both choice occasion-specific and individual-specific random coefficients, or both individual-specific and across-individual random coefficients get introduced in the model, the MSL becomes extremely time consuming and close to being infeasible to estimate (as we will show in Section 2.3.2.3). One other problem we found even in this simple panel MSL estimation was that one of the ten runs experienced non-convergence problems. On the other hand, no convergence issues were encountered with the MACML estimation.

The Non-Diagonal Case

The results for the panel non-diagonal random coefficients case are provided in Table 2.2b. As can be observed, the average APB is somewhat higher in this case relative to the diagonal case, mainly due to the APBs associated with the Cholesky parameters that determine the full covariance matrix. However, the high APB is somewhat deceiving, because the estimated values of the Cholesky parameters are not too far away from the true values. But the small values of the true Cholesky parameter values tend to inflate the APB values. Also, as indicated earlier, with a limited sample size and several parameters to estimate, this is not an unexpected result. The main point to note is that the MACML continues to do a much better job in recovering parameters than the MSL, and with at least an 8-fold or so reduction in computational cost. The actual computational efficiency is much higher, but we did not go beyond 250 Halton draws in the MSL case as in the earlier cases because of the very high computation costs involved. As importantly, we did notice an increase in convergence problems with the MSL approach in the non-diagonal case, with 5 of the 10 runs getting bogged down and not going anywhere. There were no convergence issues whatsoever with the MACML approach.

2.3.2.3 Panel Intra-Individual and Inter-Individual Random Coefficients

The MSL estimation of the situation when there are both intra-individual and inter-individual random coefficients is extremely expensive from a computational standpoint, since there are two levels of random coefficients to be integrated out (see Equations (5) and (6)). This implies that if Q draws of a QMC sequence were to be used for each level

of integration, then for each of the Q draws of the outer integral, the inner integral itself will need to be evaluated using Q draws. Thus, the number of total draws becomes Q^2. If we are to use 250 draws as we have done in the cross-sectional and panel cases earlier, the total number of draws would be 62,500. Effectively, it is practically infeasible to estimate such a model accurately using the MSL technique within a reasonable amount of time. However, the model does not pose any problems for estimation using the MACML approach; the computational cost is about the same order as for the simple panel case or the simple cross-sectional case. In this section, we provide only the results for the MACML estimation. The MSL estimation with 250 draws for each integration level was taking about 3 hours per iteration. Assuming convergence in 55 iterations (which was about the average for the case for the pure panel diagonal random coefficients case), the time for the MSL estimation would be about 165 hours or about 7 days for the diagonal covariance specification. This is in contrast to about 25 minutes for the MACML estimation of the diagonal case, and 50 minutes for the MACML estimation of the non-diagonal case.

The Diagonal Case

The results for the diagonal case are presented in Table 2.3a. The mean APB in this case is 12.6%, with the individual parameter APBs varying from 8.1% to 18.5%. The APB is affected here by the number of covariance-related parameters to be estimated, as also reflected in the higher asymptotic standard errors of the parameters compared to the previous cases. But it is indeed remarkable that the MACML method does about as well as for the panel inter-individual diagonal case of Section 2.3.2.2, both in terms of recovering parameters as well as computation time. Specifically, in terms of computation time, the model takes, on average over the 100 runs (10 different runs on 10 different data samples), only about 23 minutes for convergence, which is about twice the amount of time as for the panel inter-individual case. The standard deviation of the times for convergence over the 100 runs is about 4.5 minutes. At the same time, the approximation error remains very small, at an average of about 18% of the asymptotic standard error.

Table 2.3a Evaluation of the ability to recover true parameters for the panel intra-individual and inter-individual random coefficients diagonal case

Parameter	True Value	MACML Method				
		Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Approximation Standard Error	Approximation Adjusted Asymptotic Standard Error
Mean values of the β vector						
$b1$	1.500	1.341	10.6%	0.272	0.045	0.276
$b2$	-1.000	-0.851	14.9%	0.176	0.027	0.178
$b3$	2.000	1.753	12.3%	0.350	0.057	0.355
$b4$	1.000	0.913	8.7%	0.191	0.031	0.193
$b5$	-2.000	-1.780	11.0%	0.357	0.059	0.362
Standard deviations of the β vector						
$\sigma 1$	1.000	0.842	15.8%	0.191	0.038	0.195
$\sigma 2$	1.000	0.815	18.5%	0.179	0.028	0.181
$\sigma 3$	1.000	0.865	13.5%	0.195	0.035	0.198
$\sigma 4$	1.000	0.864	13.6%	0.188	0.032	0.190
$\sigma 5$	1.000	0.877	12.3%	0.197	0.038	0.200
$\tilde{\sigma} 1$	1.000	0.919	8.1%	0.243	0.037	0.245
$\tilde{\sigma} 2$	1.000	0.917	8.3%	0.239	0.044	0.243
$\tilde{\sigma} 3$	1.000	0.819	18.1%	0.235	0.072	0.246
$\tilde{\sigma} 4$	1.000	0.856	14.4%	0.230	0.036	0.233
$\tilde{\sigma} 5$	1.000	0.909	9.1%	0.248	0.038	0.251
Overall Mean Value			12.6%	0.233	0.041	0.236
Mean Time	22.82					
Std. dev of Time	4.53					
% of Runs Converged	100%					

Table 2.3b Evaluation of the ability to recover true parameters for the panel intra-individual and inter-individual random coefficients non-diagonal case

Parameter	True Value	MACML Method				
		Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Approximation Standard Error	Approximation Adj. Asymptotic Standard Error
Mean values of the β vector						
$b1$	1.500	1.430	4.6%	0.288	0.028	0.289
$b2$	-1.000	-0.936	6.4%	0.200	0.019	0.201
$b3$	2.000	1.925	3.7%	0.383	0.037	0.385
$b4$	1.000	0.932	6.8%	0.195	0.018	0.196
$b5$	-2.000	-1.903	4.8%	0.386	0.036	0.388
Cholesky parameters characterizing the covariance matrix of the β vector						
$l11$	1.000	0.940	6.0%	0.214	0.023	0.215
$l12$	-0.500	-0.548	9.7%	0.164	0.022	0.165
$l13$	0.250	0.188	24.6%	0.143	0.026	0.145
$l14$	0.750	0.771	2.8%	0.204	0.026	0.205
$l15$	0.000	0.103	10.3%	0.142	0.026	0.145
$l22$	0.866	0.694	19.9%	0.181	0.026	0.183
$l23$	0.433	0.332	23.4%	0.195	0.039	0.199
$l24$	-0.144	-0.030	79.2%	0.170	0.031	0.172
$l25$	0.000	0.078	7.8%	0.183	0.037	0.187
$l33$	0.866	0.810	6.5%	0.209	0.034	0.212
$l34$	0.237	0.167	29.5%	0.178	0.033	0.182
$l35$	0.000	0.003	0.3%	0.193	0.042	0.197
$l44$	0.601	0.392	34.8%	0.223	0.035	0.226
$l45$	0.000	-0.168	16.8%	0.555	0.074	0.560
$l55$	1.000	0.773	22.7%	0.408	0.060	0.413

Table 2.3b (Continued) Evaluation of the ability to recover true parameters for the panel intra-individual and inter-individual random coefficients non-diagonal case

Parameter	True Value	MACML Method				
		Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute Percentage Bias	Asymptotic Standard Error	Approximation Standard Error	Approximation Adjusted Asymptotic Standard Error
Cholesky parameters characterizing the covariance matrix of the $\tilde{\beta}$ vector						
$\tilde{l}_{11}$	1.000	1.033	3.3%	0.254	0.027	0.255
$\tilde{l}_{12}$	0.000	-0.007	0.7%	0.170	0.023	0.172
$\tilde{l}_{13}$	0.000	0.043	4.3%	0.188	0.030	0.190
$\tilde{l}_{14}$	0.000	0.030	3.0%	0.177	0.023	0.178
$\tilde{l}_{15}$	0.000	-0.008	0.8%	0.183	0.027	0.185
$\tilde{l}_{22}$	1.000	0.973	2.7%	0.249	0.028	0.251
$\tilde{l}_{23}$	0.500	0.527	5.3%	0.213	0.031	0.215
$\tilde{l}_{24}$	0.500	0.415	17.0%	0.197	0.026	0.199
$\tilde{l}_{25}$	0.500	0.422	15.7%	0.198	0.029	0.200
$\tilde{l}_{33}$	0.866	0.798	7.9%	0.272	0.036	0.275
$\tilde{l}_{34}$	0.289	0.361	25.0%	0.265	0.039	0.268
$\tilde{l}_{35}$	0.289	0.390	35.2%	0.271	0.045	0.274
$\tilde{l}_{44}$	0.817	0.720	11.8%	0.305	0.033	0.306
$\tilde{l}_{45}$	0.204	0.285	39.4%	0.300	0.045	0.303
$\tilde{l}_{55}$	0.791	0.460	41.8%	0.307	0.077	0.316
Overall Mean Value			15.3%	0.239	0.034	0.241
Mean Time		46.50				
Std. dev of Time		8.83				
% of Runs Converged		100%				

Of course, the ability to recover parameters may be improved by increasing the number of permutations per individual used in the MACML estimation. In the experiments, the APB reduced from 12.6% to 10.8%, when the number of permutations was increased to 2, and further reduced to 7.7% when the number of permutations was increased to 5. However, the mean time to convergence increased from 23 minutes (for one permutation per individual) to 44 minutes (for two permutations per individual) to 117 minutes (for five permutations per individual).

The Non-Diagonal Case

This estimation involves the most number of parameters, including five mean parameters on the five independent variables, 16 covariance elements from the individual-specific covariance matrix, and another 16 covariance elements from the choice occasion-specific covariance matrix (see Table 2.3b). The mean APB here is about 15.3%, though the APB values for the mean parameters are recovered very accurately (the APB values for the mean parameters range from 3.7% to 6.8%, which is even better than the corresponding APB values for the diagonal case). The relatively high APB values for the Cholesky parameters are a result of the high number of such parameters estimated from a sample size of 500 individuals and 2500 choice occasions, as well as the small magnitudes of the true values of the Cholesky parameters. When the number of permutations per individual was increased to two, the mean APB reduced to 14.1%. The mean APB further reduced marginally to 13.1% with five permutations per individual. The time to convergence with one permutation per individual is about twice the time needed for the diagonal case with one permutation per individual, but is still only of the order of 46 minutes on average. The times to convergence for two and five permutations per individual are also about twice the corresponding times in the diagonal case.

2.4 SUMMARY AND CONCLUSIONS

Random coefficients discrete choice models are increasingly being used for unordered response multinomial choice modeling in the transportation and other fields, as a means to accommodate varying tastes across decision makers due to unobserved (to the analyst) factors. In such random coefficients models, the overall error term vector is effectively

decomposed into an independent and identically distributed (IID) component vector and another non-IID (across alternatives) vector of jointly distributed random coefficients. While various different distributions may be used for the non-IID component, it is common practice to employ a multivariate normal distribution. If the analyst uses a normal distribution for the IID portion, the result is the "mixed" multinomial probit (MMNP) model.

The MMNP model structure may be applied to both cross-sectional and panel contexts. In either case, both the model structures do not have an analytically tractable form for the choice probabilities and for the likelihood function. The approach used to estimate such models is typically based on pseudo-Monte Carlo or quasi-Monte Carlo simulation techniques to evaluate the multidimensional integrals in these models. In such an MSL estimation approach, consistency, efficiency, and asymptotic normality of the estimator is critically predicated on the condition that the number of simulation draws per individual rises faster than the square root of the number of individuals in the estimation sample. This effectively implies that the desirable asymptotic properties of the MSL estimator are obtained at the expense of computational cost. Also, the simulation noise when dealing with high dimensionalities of integration can cause convergence problems.

Bhat (2011a) recently proposed a maximum approximated composite marginal likelihood (MACML) method for the estimation of MNP-based models. In this study, the focus is on evaluating the ability of the MACML method to recover parameters of MMNP models from finite samples, and to compare the performance of the MACML estimator with the MSL estimator in terms of finite sample bias in parameters and the computational time for estimation. Within the class of MMNP models, we examine three different model structures: the cross-sectional random coefficients structure, the panel inter-individual random coefficients structure, and the panel intra- and inter-individual random coefficients structure. Within each of these structures, both the cases of independent random coefficients (*i.e.*, the diagonal covariance specification) and dependent random coefficients (*i.e.*, the non-diagonal covariance specification) are considered.

The results of the analysis indicate that the MACML recovers parameters much more accurately than the MSL approach in all model structures and covariance specifications. The MACML inference approach also estimates the parameters efficiently, with the asymptotic standard errors being, in general, only a small proportion of the true values. It is remarkable that the approximation error involved in the use of even only a single permutation to evaluate the MVNCD function is very small, being only about 26-42% of the simulation error associated with 250 randomized and scrambled Halton draws in the cross-sectional model estimations (the simulation errors were not estimated in the MSL approach for the panel cases because of the computational costs involved in running multiple runs on the same data set). As importantly, the MACML inference approach takes only a small fraction of the time needed for MSL estimation. In particular, the results suggest that the MACML approach is about 50 times faster than the MSL for the cross-sectional random coefficients case, at least 15 times faster than the MSL for the panel inter-individual random coefficients case, and about 350 times or more faster than the MSL for the panel intra- and inter-individual random coefficients case. As the number of alternatives in the unordered-response model increases, one can expect even higher computational efficiency factors for the MACML over the MSL approach. Further, as evident in the panel intra- and inter-individual random coefficients case, the MSL is all but practically infeasible when the mixing structure leads to an explosion in the dimensionality of integration in the likelihood function, but these situations are handled with ease in the MACML approach.

Of course, the results above are specific to the mixed multinomial probit (MMNP) structure. If one insists on using the mixed multinomial logit structure (MMNL) structure, the MACML method needs to be supplemented with a normal scale mixture technique to approximate the IID extreme value error distribution. Bhat (2011b) provides the procedure. Comparisons of the MACML and MSL estimations for the MMNL model structure would be of interest for the cross-sectional and inter-individual panel cases, though the MACML procedure should continue to provide substantial computational benefits in the more involved panel intra- and inter-individual case. In any case, the

movement between the use of the MMNL and the MMNP structures has been dictated primarily by ease of estimation. In the past several years, the scale has been tilted more toward the use of the MMNL structure, primarily because of the ease of conceptualization and coding of the simulation procedure for the MMNL structure (see Bhat *et al*., 2008 and Train, 2009). However, in the MACML estimation technique, the MMNP model is easier to estimate because of the conjugate property of addition of the normal distribution. Thus, we may expect to see the scale tilting back toward the MMNP structure for the specific case when the mixing distribution is continuous and normal. When the mixing distribution is not normal, one can still use the MACML approach by approximating the continuous multivariate distribution using a multivariate normal scale mixture, but this will increase computational cost in proportion to the number of dimensions in the mixing distribution. In such cases, the MSL inference approach may continue to be the choice method for estimation of the simple models, though the MACML supplemented by the scale mixture technique would perhaps still be a promising way to proceed for more complicated models (such as the case with panel intra- and inter-individual random coefficients).

In closing, the MACML inference approach has the potential to dramatically influence the use of the mixed multinomial probit model in practice, and should facilitate the practical application of rich model structures for unordered-response discrete choice modeling.

CHAPTER 3: A MODEL OF CHILDREN'S SCHOOL TRAVEL MODE CHOICE BEHAVIOR ACCOUNTING FOR SPATIAL AND SOCIAL INTERACTION EFFECTS

The material in this chapter is drawn substantially from the following published paper.

Sidharthan, R., C.R. Bhat, R.M. Pendyala, and K.G. Goulias (2011), Model for Children's School Travel Mode Choice: Accounting for Effects of Spatial and Social Interaction. *Transportation Research Record* 2213, 78-86.[5]

3.1 INTRODUCTION

Much attention is being paid to the analysis of factors contributing to the travel mode choice behavior of children for the trip to and from school (Beck and Greenspan, 2008). Major programs aimed at promoting walking and bicycling to school are in place, particularly in the United States, where a steady decline in the shares of walk and bicycle modes for school trips has been observed over the past few decades (McDonald, 2007; McMillan, 2007; Black *et al.*, 2001). Examples of these programs include the US Department of Transportation Safe Routes to School program (http://www.saferoutesinfo.org) and the Walking School Bus initiative (http://www.walkingschoolbus.org). Much of this interest stems from the desire to promote active transportation mode use among children with a view that the choice of such modes would substantially help fight childhood obesity, which has become a serious public health concern in the United States and elsewhere (Koplan *et al*., 2005). Several studies have shown that children who use active modes of transportation for the trip to and from school are likely to be more physically active during other periods of the day as well, thus increasing the overall physical and mental well-being of children (Cooper *et al.*, 2003; Loucaides and Jago, 2008).

There are undoubtedly many factors that impact the choice of mode for the children's trips to and from school. Studies of children's school mode choice show the

[5] The coauthors provided the data for the analysis and guidance on the methodological aspects of the paper.

important effects of home-school proximity, household socio-economic attributes, neighborhood built environment characteristics, and parental or caregiver perceptions of neighborhood safety and vehicular traffic conditions on the path to and from school. A systematic review of the literature on this topic is provided by Pont *et al.* (2009); some of the pertinent literature in this topic area is reviewed in more detail in the next section of this chapter.

What is found in the literature review is that many studies loosely acknowledge, but largely ignore or do not adequately account for, the spatial interaction effects that affect children's mode choice to and from school. Spatial interaction may occur in two possible ways – across spatial units (zones, neighborhoods, tracts, blocks) because units that are closer to one another share some common unobserved attributes, and/or across behavioral units (individuals, households) because behavioral units that are closer to one another in space may share common unobserved attributes that affect the way they behave. In the context of children's school mode choice, a household's mode choice decision related to child school trips may also be influenced by the actions and choices of other households and individuals in the same spatial cluster (say, a neighborhood). For example, if parents find that many other children in the neighborhood walk to school, they may feel comfortable sending their own child by walk as well. The Walking School Bus initiative is, in fact, founded on this principle of social interaction effects among households that are in close proximity of one another.

Essentially spatial interaction among individuals may arise in the context of children's mode choice to school in a number of ways. Similarly, social interactions among parents in a neighborhood or whose children attend the same school could lead to exchange of information about characteristics of different modes thus contributing to a dependence in the mode utility functions of different individuals. Another possible way in which such correlation can arise is one where other children in the same neighborhood using an active mode of transportation create a positive environment for the use of such modes by improving the safety of walking/bicycling in the neighborhood, and this might persuade other children and their caregivers to adopt non-motorized modes of

transportation for the trip to and from school. Finally, similarities in the built environment attributes across households/individuals who are located in greater proximity of one another may also create interactions in the modal utilities of individuals.

Previous attempts to study school mode choice for children have not accounted for such spatial and social interaction effects, although some attempts have been made to consider spatial attributes in mode choice decisions (*e.g.*, Mitra *et al.*, 2010). The accounting for such effects requires methodological advancements in the specification and estimation of discrete choice models; this study is aimed at presenting a methodological framework and estimation approach that makes it possible to estimate mode choice models with spatial and social interaction effects. Another major impediment to the development of mode choice models that account for spatial effects is that detailed spatial accessibility measures at small levels of geography are generally not available in most travel survey data sets. In this particular study, disaggregate census tract-level spatial accessibility measures are computed based on Chen et al. (2010) for a survey sample drawn from the 2009 US National Household Travel Survey (NHTS) and used in the study to disentangle unobserved spatial correlation effects from observable built environment attributes associated with household location.

Following this brief introduction, an overview of the literature is offered in the next section. The section 3.3 presents the modeling methodology adopted in the study. The section 3.4 provides a description of the data set while the section 3.5 summarizes model estimation results and study findings. The final section offers concluding thoughts and directions for further research.

3.2 ANALYSIS OF CHILDREN'S SCHOOL MODE CHOICE

There has been considerable amount of research aimed at studying children's school trip mode choice behavior. Pont *et al.* (2009) provides a systematic review of the literature on this topic and more broadly on the topic of active transportation among children. Studies on children's school mode choice span the globe as this is clearly an issue of interest in metropolitan contexts around the world. In the US, an analysis by McDonald (2007) of the series of national travel surveys from 1969 through 2001 shows the

substantial decline in active mode use over the past several decades. In 1969, about 41 percent of students bicycled or walked to school; by 2001, that proportion had decreased to just about 13 percent. McDonald (2007) indicates that the increase in distance between home and school may account for about one-half of the decline in the use of active transportation modes to school.

Distance between home and school is a critical factor affecting the use of non-motorized modes (Lawrence Frank and Company, Inc., 2008). Ewing *et al.* (2004) analyzed data from Gainesville (Florida) and found distance to be one of the most important factors in the choice of bicycle and walking modes. Yeung *et al.* (2008) report a similar result in an analysis of data from Brisbane, Australia. However, unlike the US study, they did not find any significant difference in the body mass index (BMI) of children using active modes of transport versus those using motorized modes for travel to and from school. Loucaides and Jago (2008), analyzing data from Cyprus, find that overweight children who walked to school were more physically active in general when compared with overweight children who were driven to school. However, no such difference was observed across normal weight children. Cooper *et al.* (2003) analyze a sample from Bristol, UK, and report that boys who walk to school are likely to be more physically active in general after school than those who used motorized modes of transport. Such differences were not found among girls.

There are several studies dedicated to analyzing the influence of the built environment attributes and street configuration on school mode choice. The results are somewhat mixed, possibly due to the difficulty in measuring built environment attributes and appending such variables to individual person and household survey records. For example, Yarlagadda and Srinivasan (2008) found strong impacts of socio-economic attributes and distance, but report that the impacts of travel time and built environment attributes are statistically insignificant. Similarly, McMillan (2007) reports that urban form variables had a modest impact on mode choice; these variables had a relatively less impact than other variables representing socio-economic attributes, distance, and vehicular traffic conditions. On the other hand, Boarnet *et al.* (2005), in analyzing the

impact of the Safe Routes to School program, found that sidewalk improvement, crossing improvements, and traffic control enhancements improved the odds of children switching to walk and bicycle modes. Ewing *et al.* (2004) also note that street density and sidewalk connectivity are influential in facilitating walking to school.

Traffic safety and parental perception of crime against children (*e.g.*, abduction, molestation) were found to be significant in a few studies. Timperio *et al.* (2006), in an analysis of data from Melbourne, Australia, found that parental perception of the number of children walking to school in the neighborhood, presence of lights and adequate crossings, and the presence of a busy roadway between the home and school impacted school mode choice. DiGuiseppi *et al.* (1998), in a study of data from the UK find that adults accompanied 84 percent of children to and from school. Only three percent of bicycle users were allowed to bicycle on main roads. Ninety percent of parents were very or quite worried about abduction or molestation and an almost identical percentage were very or quite worried about traffic safety. They found parental concerns about safety were strong predictors of school mode choice.

Some studies have identified a few other factors influencing school mode choice. Weather conditions are cited as an important explanatory variable by Muller *et al.* (2008) in a study conducted in Germany, while psychological and attitudinal factors are found to be significant by Black *et al.* (2001) who report on a study conducted using data gathered from 51 schools in the UK. Zwerts *et al.* (2009), in a study of Belgian students, find that students viewed the walking and bicycling experience en route to school as an important factor in the attractiveness of those modes. Dellinger and Staunton (2002) analyzed data from the US National Health survey (conducted by the Centers for Disease Control and Prevention). They report that barriers to walking and bicycling among children were long distances, traffic danger, and adverse weather conditions. They find that 85 percent of those who reported no barriers ended up using non-motorized modes of transportation.

The role of parental influence, intra-household interactions, and social networks is further brought out in other studies. For example, the study by Yarlagadda and Srinivasan (2008) explicitly focuses on the escort person for the school trip. They report

that the presence of multiple school-going children in the household increases the odds that the mother will drive the children to school. This finding is in contrast to that reported by McDonald (2008), who notes that having siblings increases the likelihood of walking and reduces the likelihood of being driven. These findings point to the need to further study the role of intra-household interactions in school mode choice behavior. McMillan *et al.* (2006) found that the odds of biking or walking to school are 40 percent lower in girls than boys, but note that the relationship is moderated by the caregiver's own walking propensity and behavior. Pooley *et al.* (2010) examined GPS traces of school journeys of children in the UK and find great variability in the characteristics of school travel. They attribute this variability to complex household interactions, family responsibilities, personal commitments, and personal preferences. Zwerts *et al.* (2009) note that the social aspect associated with walking or bicycling together is very important, particularly for girls.

From the review of the literature, it is clear that several factors influence school mode choice for children. While some results are mixed, it is clear that home-to-school distance (proximity), socio-economic characteristics, built environment attributes, street configuration, land use density and mix, and attitudes and perceptions of safety and crime are important determinants of school mode choice behavior. While these studies acknowledge the potential importance of interactions within and outside the household arising from neighborhood effects, and a couple of studies attribute certain results obtained to intra-household interactions and neighborhood social networks, the studies do not explicitly account for interaction/social network effects in the modeling of school mode choice. Mitra *et al.* (2010) analyze data from Toronto and use spatial autocorrelation measures to identify zones with high walking rates. However, their study does not involve the estimation of a mode choice model in the presence of spatial interaction effects. Ulfarsson and Shankar (2008) also attempt to capture correlation effects, but the focus of their model specification is on accounting for correlations across alternatives using a covariance heterogeneity specification (as opposed to capturing interaction effects across behavioral units over space).

This study aims to fill a critical gap in the understanding of children's school mode choice behavior by developing a model that accounts for spatial and social effects arising from interactions among household members and across households in geographical and social clusters, respectively.

3.3 MODELING METHODOLOGY

Spatial interaction effects may exist across discrete choice alternatives (*e.g.*, Bolduc *et al.*, 1996; Bhat and Guo, 2004) or across decision-makers (*e.g.*, Anselin, 2003; Bhat and Sener, 2009). The focus in this study is on spatial and social interactions across decision makers. Interestingly, in the context of spatial interaction across decision makers, earlier studies have either focused on binary response models or ordered response models. In particular, spatial interaction across individuals has seldom ever been discussed in the context of unordered-response models. However, spatial interaction in data may occur in unordered-response models for the same reasons (for example, diffusion effects, social spillover effects, and unobserved location-related effects) that these effects have been studied extensively in binary and ordered-response models.

In terms of estimation of binary and ordered-response discrete choice models with a general spatial structure, the analyst confronts, in the familiar probit model, a multi-dimensional integral over a multivariate normal distribution, which is of the order of the number of observational units in the data. While a number of approaches have been proposed to tackle this enormous multidimensional integration problem (*e.g.*, LeSage, 2000; Fleming, 2004), none of these methods are practically feasible for moderate-to-large samples as they are quite cumbersome from a computational standpoint. In the context of unordered-response models, the situation becomes even more difficult – the likelihood function entails a multidimensional integral over a multivariate normal distribution of the order of the number of observational units factored up by the number of alternatives minus one. This situation, however, is relatively easily handled using the Maximum Approximated Composite Marginal Likelihood (MACML) estimation method proposed by Bhat (2011a).

3.3.1 Model Formulation

Consider a spatial lag model structure for unordered-response models as proposed by Bhat (2011a), where the dependencies in modal utilities across individuals is caused by a combination of direct spillover effects (utilities of individuals "rubbing" off on each other) and indirect unobserved spatial/social effects. In such a model structure, the utility that an individual q associates with alternative i (i = 1, 2, …, I) is assumed to take the following form:

$$U_{qi} = \rho \sum_{q'} w_{qq'} U_{q'i} + b' x_{qi} + \xi_{qi};\ \xi_{qi} \sim N(0,0.5), |\rho| < 1, \quad (1)$$

In the above formulation x_{qi} is a (K × 1)-column vector of exogenous attributes, b is a (K × 1)-column vector of corresponding coefficients, $w_{qq'}$ is the spatial weight corresponding to individuals q and q', with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$ for each (and all) q. It is also assumed that ξ_{qi} is independent and identically distributed across q and i. The above utility function may be equivalently written as:

$$U_{qi} = \left[(IDEN_Q - \rho W)^{-1} x_i b\right]_q + \left[(IDEN_Q - \rho W)^{-1} \xi_i\right]_q, \quad (2)$$

where $IDEN_Q$ is an identity matrix of size Q, W is a spatial weight matrix of size $Q \times Q$, x_i is a $Q \times K$ vector $(x'_{1i}, x'_{2i}, ..., x'_{Qi})$, ξ_i is a vertically stacked vector of the ξ_{qi} terms of size $Q \times 1$, and $[.]_q$ indicates the q^{th} element of the column vector [.]. Substituting $V_{qi} = \left[(IDEN_Q - \rho W)^{-1} x_i b\right]_q$ and $\varepsilon_{qi} = \left[(IDEN_Q - \rho W)^{-1} \xi_i\right]$, Equation (2) may be written equivalently as,

$$U_{qi} = V_{qi} + \varepsilon_{qi} \quad (3)$$

where $\text{Var}(\varepsilon_i) = \tilde{A} = 0.5(IDEN_Q - \rho W)^{-1}(IDEN_Q - \rho W')^{-1}$, ε_i being the vertically stacked vector of the Q ε_{qi} error terms. Define $H_{qim_q} = V_{qi} - V_{qm_q}$, where m_q is the alternative chosen by individual q.

Then, the latent utility differentials, $y^*_{qim_q}$ $(= U_{qi} - U_{qm_q}, i \neq m_q)$, may be written as: $y^*_{qim_q} = H_{qim_q} + (\varepsilon_{qi} - \varepsilon_{qm_q})$, $i \neq m_q$. Let $y^*_q = (y^*_{q1m_q}, y^*_{q2m_q}, ..., y^*_{qIm_q};\ i \neq m_q)'$, and let $\boldsymbol{y}^* = (y^{*'}_1, y^{*'}_2, ..., y^{*'}_Q)'$. Thus $\boldsymbol{y}^*$ is an $(I-1)*Q$ vector. Also, let $H_q = (H_{q1m_q}, H_{q2m_q}, ..., H_{qIm_q};\ i \neq m_q)$, which is an $(I-1)\times K$ matrix. The likelihood of the observed sample (*i.e.*, individual 1 choosing alternative m_1, individual 2 choosing alternative m_2, …, individual *Q* choosing alternative m_Q) may then be written succinctly as Prob[$\boldsymbol{y}^*$ < **0**]. To write this likelihood function, note that the mean vector of $\boldsymbol{y}^*$ is

$$B = \left[(H_1)', (H_2)', ..., (H_Q)'\right]'. \tag{4}$$

Then we can write $\boldsymbol{y}^* \sim MVN(B, \Sigma)$, and the likelihood function of the sample is:

$$L_{ML}(b, \rho) = \text{Prob}(y^* < 0) = F_{(I-1)*Q}(-B, \Sigma). \tag{5}$$

where Σ covariance matrix of $\boldsymbol{y}^*$ and Bhat (2011a) provides the equations for calculating this. $F_{(I-1)*Q}$ is the multivariate cumulative normal distribution of $(I-1)*Q$ dimensions. Of course, maximizing the above likelihood function requires the evaluation of an $(I-1)*Q$ integral. Integrals of this high dimensionality are clearly impractical to evaluate using the usual Monte Carlo simulation methods. However, the MACML estimation approach can be used here. In the MAMCL estimation approach, a combination of the composite marginal likelihood method and the approximation method for multivariate normal orthant probabilities is used. The pairwise CML function for the sample is given by the expression below:

$$\begin{aligned} L_{CML}(b, \Omega, \rho) &= \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \text{Prob}\,(C_q = m_q, C_{q'} = m_{q'}) \\ &= \prod_{q=1}^{Q-1} \prod_{q'=q+1}^{Q} \text{Prob}[\, y^*_{qim_q} < 0 \forall i \neq m_q \text{ and } y^*_{qim_{q'}} < 0 \forall i \neq m_{q'}\,] \end{aligned} \tag{6}$$

Each multivariate orthant probability above has a dimension equal to $(I-1)\times 2$, which can be computed using the approximation proposed by Bhat (2010) in the MACML Approach. The variances and correlations in the bivariate and univariate cumulative

normal distribution expressions in the approximation can be obtained as appropriate sub-matrices of Σ. An issue that has a direct impact on computational time in the CML approach is the number of pairs ($=Q(Q-1)/2$ pairs) of $(I-1)\times 2$ multivariate probability computations.

The framework discussed above is extendable to include social and other forms of dependence as well. This is because the weight matrix W that forms the basis for spatial dependence can also be the basis for more general forms of dependence. In fact, W itself can be parameterized as a finite mixture of several weight matrices (as in Yang and Allenby's (2003) application to the simple binary choice model), each weight matrix being related to a specific covariate *k*, *i.e.*, $W = \sum_{k=1}^{K} \varphi_k W_k$, where φ_k is the weight on the k^{th} covariate in determining dependency between individuals ($\sum_{k=1}^{K} \varphi_k = 1$), and W_k is a measure of distance between individuals on the k^{th} covariate.

3.4 DATA

The data used in this study is derived from the California add-on sample of the 2009 National Household Travel Survey (NHTS) conducted in 2008-2009 in the United States. Within the California add-on sample, the survey subsample of respondents from the Los Angeles – Riverside – Orange County region was extracted and used for the model estimation effort. This selection process was done for several reasons. First, the use of a national sample for studying school mode choice behavior may be inappropriate given that there are likely to be substantive geographic differences across the country. Spatial correlation effects are likely to be more localized in nature, calling for the use of data drawn from a more limited geographic region for analysis and model development. Second, the use of a very large sample for model estimation would produce inflated test statistics that would affect inferences drawn from the model results. Finally, the authors have access to census tract-level accessibility measures and land use data for the Los Angeles region in conjunction with an ongoing activity-based model development effort underway for the Southern California Association of Governments (SCAG).

The survey collects detailed socio-economic, demographic, and travel information for all household members in respondent households. The survey also collects information about usual travel characteristics by asking questions about travel undertaken in the past week. Extensive descriptive statistical analysis was conducted on the data to understand mode choice patterns for children's school trips and to identify explanatory factors that may influence such behavior. For the sake of brevity, all of the analysis conducted is not described and presented here, but some highlights are noted to provide an overview of the data assembly process in a nutshell.

Table 3.1 Sample Demographic Characteristics

Characteristic	Value
Average travel time to school (min)	12.4
Average travel time to school by modal market segment (min)	
Car	10.9
School Bus	25.8
Bicycle	14.0
Walk	12.1
Car – School Bus	16.7
Car – Walk	9.7
Median household income	$70,800
Median household income by modal market segment (min)	
Car	$78,000
School Bus	$50,300
Bicycle	$73,600
Walk	$54,700
Car – School Bus	$82,700
Car – Walk	$68,400
Number of household members	4.3
Number of vehicles in household	2.4
Number of bicycle trips in past week	1.3
Number of walk trips in past week	4.0
Number of adults in household	2.3
Number of workers in household	1.5

The survey sample included 1192 children aged 5-15 years for whom school mode choice behavior could be analyzed. Table 3.1 presents the overall average travel time to school,

the average travel time by mode used, the overall median household income value, the median household income by mode used, and brief descriptive statistics of other household characteristics to which these children belong. In general, the travel time to school ranges from about 10 minutes to 15 minutes with an overall average of 12.4 minutes. Only the average bus travel time falls outside this range with an average value of just over 25 minutes. Those who walk and use the school bus report lower median household incomes than other groups. Thus, it is clear that mode choice to and from school is correlated with income; perhaps the lower car ownership in these households lead children to walk and use the school bus. In general, the household characteristics show that households are larger than would be expected if one were analyzing the general population. This is consistent with the fact that the analysis sample here is exclusively focusing on households with children going to school.

Table 3.2 School Mode Choice Distribution by Distance from Home to School

Mode	Distance home to school					Total
	< ¼ mile	¼ - ½ mile	½ - 1 mile	1-2 miles	> 2 miles	
Car	23.9%	46.2%	56.8%	68.5%	74.7%	44.3%
School Bus	-	2.5%	1.2%	7.6%	15.7%	33.8%
Bicycle	2.5%	1.7%	4.9%	2.0%	0.5%	1.2%
Walk	60.4%	37.0%	22.2%	10.4%	1.1%	7.0%
Car-School Bus	-	-	0.6%	0.8%	6.4%	10.1%
Car-Walk	13.2%	12.6%	14.2%	10.8%	1.6%	3.5%
Total children	159	119	162	251	439	1130
% by distance	14.1%	10.5%	14.3%	22.2%	38.8%	100%

The importance of distance in school mode choice behavior has been highlighted in previous research. Table 3.2 presents modal split distributions by home-to-school distance bands. The association between home-school distance and modal split is readily apparent. While the overall mode split for car is 44 percent, the highest among all modes, it is clear that walk is the predominant choice of mode at very short distances. At distances less than a quarter-mile, 60 percent of children walk to school and less than

one-quarter take the car to school. However, 13 percent of children use a combination of car and walk (*i.e.*, they take the car to school, but walk back home after school). There is a dramatic increase in car mode share as distance increases; the car mode share nearly doubles to 46 percent at distances over a quarter-mile but under a half-mile. The car mode share continues to increase with distance and reaches nearly 75 percent at home-to-school distances in excess of two miles. The school bus mode share also increases with distance, consistent with expectations. The bicycle mode share shows some fluctuations, with higher shares seen for very short trips under a quarter-mile, and mid-range distances of one-half to two miles. The car-school bus combination shows a significant modal percent (6 percent) at longer distances, again consistent with expectations. Walk mode share dramatically drop off with increasing distances, with just about a one percent mode share for school trip distances greater than two miles. One of the factors affecting the choice of active modes of transportation is that nearly 40 percent of the children live more than two miles away from their school. Only about 25 percent of the children live within a half-mile of their school location. As schools get increasingly larger and cover larger boundary areas, this challenge may become more pronounced.

An analysis of the data showed that some children use a combination of modes to commute to and from school. In a cross-classification table of modes to and from school (the table is not presented here due to space considerations), the diagonal elements of the table show the largest figures as expected, signifying that the vast majority use the same mode to and from school. Of the 1192 children, 1041 (87 percent) use the same mode to and from school. More than one-half of the children use the car in both directions, while close to 20 percent walk in both directions. Among the modal transition segments, the largest one (with 71 students) involves the use of the car to go to school and walking back home from school. Other modal transitions are rather small, although the walk-car and car-school bus segments cannot be ignored.

In preparing the final data set for model estimation, modes with very few observations were eliminated. These included “other”, “school bus + walk”, and “bicycle + car”. This left 1143 students in the children sample. After further cleaning the data set,

removing observations with missing information and clearly miscoded values, and other reductions, 800 observations were retained for estimation.

In the survey, the walk travel time was reported for those children who walked to school. In addition, the distance between home and school was obtained for all the children in five distance bands (see Table 3.2). In examining the walk travel times and the distances to school for children who actually walked, we found that there was a good bit of variation in walk times within the sample of children who were in the same distance band. So, we decided, from an econometric efficiency perspective, to consider both travel time and travel distance in the specification. In doing so, we imputed the walk time to school for those children who did not walk to school by computing the mean walk travel time for children who do walk to school in the corresponding distance band. However, as reported later, walk travel distance did not turn out to be significant after controlling for walk travel time. For other modes, we similarly developed imputation procedures to construct travel time values for all individuals (whether or not they used the mode), and considered both travel times and distances (in the five distance bands). Interestingly, for all the non-walk modes, the distance variable specification turned out to be better, presumably because of rounding and inaccuracy in trip time reporting for these relatively long trips.

As mentioned earlier, there may be household interactions that affect choice of mode for school trips. The bicycling and walking activity of adults in each household is reported in the survey as the number of bicycling and walking trips undertaken for various purposes in the previous week. For this study, adults (parents) were classified as active bicyclists or walkers if they made at least five trips using the corresponding mode in the previous week, with at least one trip being made for a purpose other than to escort children to and from home. In other words, if the sole reason that an adult made bicycle or walk trips in the past week is to escort children, then the person is not considered an active bicyclist or walker (to avoid potential endogeneity problems).

The NHTS data set includes a set of attitudinal variables that capture individual attitudes and perceptions. In particular, the survey asks parents to rate a series of issues

on a five-point scale with one meaning that the particular consideration is not an issue and five meaning that the consideration is a serious issue. Adults were asked to identify the extent to which each of the following considerations affected the decision to allow their child (children) to walk or bicycle between home and school: distance between home and school, amount of traffic along the route, the speed of traffic along the route, the violence or crime along the route, and poor weather or climate in the area. A principal components factor analysis (without rotation) was undertaken to reduce these five attitudinal variables into a set of orthogonal factors. The factor analysis yielded two factors, one corresponding to objectively measurable attributes such as distance, and speed and volume of traffic, and the other corresponding to more subjective measures of crime and weather. These factors were used in the model specification to capture effects of parental attitudes on school mode choice.

3.5 MODEL ESTIMATION RESULTS

A simple probit model that does not account for spatial/social interaction effects and the spatial interaction model were estimated, and the estimation results are presented in Table 3.3. A systematic procedure in which variables were entered in a stepwise manner and checked for their statistical significance and intuitive behavioral interpretation was followed to arrive at the final model specification. Various forms of explanatory variables and interaction effects among them were tested to arrive at the best possible model specification that is parsimonious, and yet sensitive to a range of effects that one would expect to see in a mode choice model of the type developed in this study. An examination of the alternative specific constants shows that, in general, the bicycle and car+walk combination modes are generally less preferred than other modes (though the constants also control for the range of exogenous variable values in the sample). It is also found that there are substantial differences in the alternative specific constants between the probit model with no spatial/social effects and the spatial interaction model. This is a first indication that ignoring spatial interaction effects, when in fact they are present, results in inaccurate estimates of preferences for alternative modes. With respect to travel characteristics, findings are largely consistent with expectations. As the time to

walk increases, the utility of walking decreases. For distances less than two miles, the utility of school bus decreases; presumably the bus is of greater value when distances to school are more than two miles. On the other hand, the utility of bicycle and car+walk combination modes is higher for distances within this range.

Table 3.3 Model Estimation Results

Variable Category	Variable	Mode Utility Equation	Independent Probit		Spatial Model	
			Coef	Est/std err	Coef	Est/std err
Alternative Specific constant	School Bus		0.393	1.864	1.413	4.819
	Bicycle		-3.721	-4.724	-1.942	-2.845
	Walk		0.570	2.112	1.420	4.307
	Car + School Bus		-0.858	-5.424	1.133	0.869
	Car + Walk		-2.000	-6.814	-1.103	-3.837
Trip Characteristics	Time to walk	Walk	-0.061	-11.522	-0.063	-10.599
	Dist. to school < 2 miles	School Bus	-0.542	-3.740	-0.521	-3.635
	Dist. to school < 2 miles	Bicycle, Car+Walk	1.099	6.458	1.123	5.908
	Dist. to school < 2 miles	Car+School Bus	-1.629	-0.285	-1.725	-0.724
Individual Demographics	Age	Bicycle	0.157	2.537	0.165	2.575
	Age	Walk, Car+Walk	0.056	3.236	0.058	3.298
	Female child	Bicycle	-0.920	-2.164	-0.943	-2.519
Household Demographics	Household Income	School Bus	-0.107	-5.270	-0.103	-5.244
	Household Income	Walk	-0.065	-3.989	-0.063	-3.971
	Vehicles per capita in household	Car	0.440	2.545	0.447	2.514
	Adult non-worker present in household	Walk	0.301	2.186	0.318	2.194
	Adult non-worker present and #cars > #workers	Car	0.308	3.243	0.313	3.214
Parents Attitude	Attitude towards walk/bicycle mode	Walk	-0.102	-1.919	-0.108	-1.953
Accessibility of neighborhood	Total amount of retail employment that can be reached in 10 minutes	School Bus	-0.055	-2.049	-0.044	-2.160
Spatial Interaction parameter	ρ		-	-	0.844	6.447
	CML		-584880.8		-580600.8	

Age and gender of the student are found to be statistically significantly associated with school mode choice. The utility of bicycling, walking, or using a combination of car and walk increases with the age of the child. In other words, older children are more likely to use non-motorized modes of transportation than younger children, presumably because parents feel more comfortable letting older children use these modes. It is interesting to note that the coefficient associated with age is substantially higher for the bicycle mode than for the walk modes, suggesting that the utility for bicycle increases more than for walk modes with increasing age. A gender effect is apparent with females less likely to choose the bicycle than their male counterparts, a finding previously reported by McMillan *et al.* (2006).

With respect to household demographics, higher household income and vehicle ownership is associated with greater propensity to use the car and lower utility for alternative modes – school bus and walk. This is consistent with previous research that also reports that households with higher levels of vehicle ownership are less likely to depend on alternative modes for transporting children to and from school (McDonald, 2008; Ewing *et al*., 2004). The presence of adult non-workers in the household positively impacts the use of the walk mode, perhaps because the adult non-worker can accompany the child on the walk to and from school (alleviating safety concerns associated with having the child walk alone). However, when there are one or more adult non-workers in the household with a spare automobile, then the utility of car increases. The parental attitude is captured through the attitudinal factor that measures whether the parents considered distance and traffic conditions to be issues associated with having their child(ren) commute by walk or bicycle. If the attitudinal factor value increases, then it means that the parents considered the issue to be more serious. As expected, in households where parents had issues with distance and traffic conditions, the utility of walk for commuting to and from school decreases. Interestingly, the subjective attitudinal factor (capturing weather and safety concerns) was not statistically significant. Physically active parent, either by being active bicyclist or active walker, increases the probability of child using the corresponding modes themselves. However the relationship

seems to be weak and the coefficients were insignificant at 0.05 level of significance. So these parameters are not included in the final results presented in this study.

Spatial factors play an important role in determining school mode choice. The accessibility of the neighborhood is measured by the total amount of retail employment that can be reached within a 10-minute radius of the home location. These accessibility measures were computed at the tract-level using block-level data about employment in different industry sectors obtained from the Southern California Association of Governments (SCAG). In general, it is found that a higher level of neighborhood accessibility (measured in terms of retail employment) has a negative association with school bus mode utility. It is possible that these households are in higher density areas more conducive to walking and bicycling, or have busy streets that motivate the use of the car. This finding is consistent with that reported previously by Ulfarsson and Shankar (2008), Yarlagadda and Srinivasan (2008), and Ewing *et al.* (2004).

Spatial interaction effects were tested by specifying the weight matrix using both geographical proximity and demographic closeness as potential measures of the correlation. For geographic proximity, alternate specifications of distance (*e.g.*, inverse of distance between individuals, inverse of exponentiated distance) and membership in a county (w_{ij} = 1 if i and j belong to the same county; w_{ij} = 0 otherwise) were used. The distance between individuals was obtained as the distance separation between the centroids of the tracts of the household locations of individuals. For demographic closeness, alternate specifications of income and age similarity were created using demographic distance measures. For each of these specifications, parameters were estimated independently using the MACML approach described in this study. The social interaction effects turned out to be statistically insignificant in all demographic distance-based weight matrix specifications. The spatial interaction parameter was significant (and positive) for all geographic distance-based weight matrix specifications and the best CML was obtained for the specification using the inverse of distance as the spatial proximity measure.

The spatial correlation parameter ρ is positive, high in magnitude (0.844), and statistically significant indicating that there is high degree of geographical interdependence in the choice of mode of travel to school. This indicates that the spatial lag model is more appropriate than the non-spatial independent multinomial probit (IMNP) model. Another way to demonstrate this is to use the adjusted composite maximum likelihood ratio test (ADCLRT) statistic, which follows a chi-squared distribution (see Pace *et al*., 2011, Bhat, 2011a). This statistic returned a value of 17.2 for comparing the spatial lag model with the IMNP model, which is higher than the corresponding chi-squared table value with one degree of freedom at any reasonable level of significance. However, and very importantly, the difference between the IMNP model and the spatial lag model is not simply a matter of data fit. The effects of a change in variable on aggregate mode shares will be quite different between the two models, because the IMNP model ignores interdependence, while the spatial lag model accommodates spillover effects due to interactions between decision-agents and so may lead to relatively large changes in aggregate mode behavior despite only small changes in the underlying primitives (or determinants) of the behavior. To demonstrate this difference in effects between the IMNP and the spatial lag model, we examined the effect of a 5% decrease in walk time to school (say due to better siting of schools relative to residences) and the impact of a 25% decrease in the level of negativity in parental attitude (in the context of distance and traffic conditions being deterrents) toward allowing children to travel to school by walk or bicycle. The decrease in walk time is estimated to lead to a 0.29% decrease in car mode share according to the IMNP model, but a decrease in car mode share by almost 12% according to the spatial lag model. Similarly, the improvement in parental attitude toward non-motorized modes is estimated to decrease the car mode share by just 0.48% according to the probit model, but by 3.2% as per the spatial lag model. Clearly, the spillover effects are at work here, and the IMNP model provides estimates that are quite different than the spatial lag model.

In summary, the spatial interdependence means that, for any individual, the utility of each alternative is positively (negatively) influenced by an increase (decrease) in the

utility of corresponding alternatives for his/her geographical neighbors. In other words, the spatial dependence in school mode choice appears to arise more from social interaction and neighborhood location effects associated with households *geographically clustered closer together*. It is possible that parents of households living in a zone or tract or neighborhood interact with one another and share experiences about school travel of their children. Households may band together to facilitate walking and bicycling in a safe and secure way, but this interaction among households is more due to geographic proximity considerations as opposed to socio-economic similarity considerations (although it is plausible that households living within a neighborhood are at least somewhat homogeneous with respect to socio-economic characteristics). When other children in the neighborhood use a mode like bicycling or walking, this creates a positive externality by improving the safety of bicycling and walking in the neighborhood, thus enhancing the utility of these modes for any particular household in the neighborhood. As households in a geographical cluster are likely to deal with the same or similar built environment, it is not surprising that the geographic distance-based spatial interaction parameter turned out to be statistically significant.

3.6 CONCLUSIONS

This research has focused on the modeling of school trip mode choice behavior among children (less than 15 years of age) with a view to examine for the presence of spatial and social interaction effects that may impact such behavior. These effects may arise due to interactions among households that are geographically or demographically similar to one another. When such interaction effects are present, the modal utilities of individuals become dependent, thus violating the basic assumption of traditional discrete choice models which assume independence of error terms across observations. The usual maximum likelihood estimation of a model that accounts for global spatial/social effects is quite complex as one must evaluate very high dimensional integrals of a multivariate normal distribution to compute the likelihood function (the order of the integral is the number of observations multiplied by the number of alternatives minus one; in the empirical context of the current study, this translates to 4000-dimensional integrals). In

this study, a maximum approximated composite marginal likelihood (MACML) approach recently developed by Bhat (2011a) is employed to estimate a school mode choice model that accounts for spatial interaction effects.

In this study, the MACML approach is applied to a sample of children in the Southern California (Los Angeles and surrounding cities) region of the United States using data collected as part of the 2009 National Household Travel Survey (NHTS). The survey sample includes 800 children who provided detailed mode choice information for the journey to and from school along with information about household member use of bicycle and walk modes, and parental concerns about the built environment in relation to their children's use of bicycle and walk for traveling to and from school. Both an independent probit model (that does not account for spatial interaction effects) and a spatial correlation model were estimated to see whether the spatial interaction effects are indeed significant and present. It is found that the spatial correlation, arising from interactions among households that are geographically clustered, is statistically significant.

The findings in this study suggest that the consideration of spatial interaction effects is important in modeling mode choice behavior, particularly in the context of children's school mode choice, where residential proximity-based interaction among households and children is likely to be prevalent. Essentially, this means that programs aimed at enhancing bicycle and walk as modes of choice for the trip to and from school (such as the Safe Routes to School program in the United States) should be focused in such a way that it maximizes the likelihood of interactions based on geographic proximity. That is, given that spatial interaction effects fade over distance (based on the inverse distance specification for the spatial weights), one can use an optimization program to define the boundaries of "fixed' neighborhoods to maximize interaction effects.

CHAPTER 4: INCORPORATING SPATIAL DYNAMICS AND TEMPORAL DEPENDENCY IN LAND USE CHANGE MODELS

The material in this chapter is drawn substantially from the following published paper.

Sidharthan, R., and C.R. Bhat (2012), Incorporating Spatial Dynamics and Temporal Dependency in Land Use Change Models. *Geographical Analysis* 44(4), 321-349[6]

4.1 INTRODUCTION

This study proposes a new econometric approach to specify and estimate a model of land-use change, based on the now rich theoretical literature on land use conversion decisions made by economic agents to maximize net returns (see Plantinga and Irwin, 2006). As such, the motivations of this study stem both from a methodological perspective as well as an empirical perspective. At a methodological level, the study focuses on specifying and estimating a multi-period multinomial probit model, accounting for observation unit-specific inter-temporal dependencies, and a spatial lag structure across observation units. The model also accommodates spatial heterogeneity in the model. The model should be applicable in a wide variety of fields where social and spatial interactions (or dyadic interactions) between decision agents lead to spillover effects. The inference methodology used is the maximum approximate composite marginal likelihood (MACML) approach proposed by Bhat (2011a), and is strongly motivated by the very difficult computational problems that arise from the use of a Bayesian Markov chain-Monte Carlo (MCMC) or classical maximum simulated likelihood (MSL) inference approaches. At an empirical level, the study models the discrete indicators for the type of land-use of each spatial unit within a discrete choice model framework. The model brings together the quantitative (but aspatial or highly stylized spatial effects) perspective of

[6] The coauthor provided guidance on the methodological aspects of the paper.

land-use analysis that dominates the economic literature with the qualitative (but richer spatial dynamics and heterogeneity) perspective of land-use analysis that is quite prevalent in the ecological literature (see Irwin, 2010 for a discussion of the different perspectives of economists and ecologists in the context of urban land use change analysis). In this manner, the current study also attempts to develop a stronger linkage between the spatial unit of analysis used in economic models of land-use change and the dyadic interactions between land-owners of proximally-spaced spatial units. Thus, the empirical model is closely tied to the underlying theoretical underpinnings of the land-use model.

The next section discusses the econometric context for the current study, while the subsequent section presents the empirical context.

4.1.1 The Econometric Context

There are four precursors of the current research that are worth noting. The recent studies by Carrión-Flores *et al*. (2009) and Smirnov (2010) superimposed a spatial lag structure over a multinomial logit (MNL) model. Carrión-Flores *et al.* estimated the resulting spatial model using a linearized version of Pinkse and Slade's (1998) Generalized Method of Moments (GMM) approach (as proposed by Klier and McMillen, 2008 for the binary choice model), while Smirnov employed a pseudo-maximum likelihood (PML) estimator to obtain model parameters. Smirnov's PML estimator is essentially based on estimating the spatial autoregressive term in the spatial lag model by recognizing the implied heteroscedasticity generated by the spatial correlation, while ignoring the spatial correlation across observational units. The approaches of Carrión-Flores *et al*. and Smirnov simplify inference by avoiding multidimensional integration. However, they are both based on a two-step instrumental variable estimation technique after linearizing around zero interdependence, and so work well only for the case of large estimation sample sizes and weak spatial dependence. Chakir and Parent (2009) estimated a multinomial probit model of land-use change, similar to the empirical focus of the current study. However, they employed a Bayesian MCMC method, which requires extensive simulation, is time-consuming, is not straightforward to implement,

and can create convergence assessment problems.[7] Sener and Bhat (2012) allowed spatial error dependence in a multinomial logit model of choice, but their approach is not applicable to a spatial lag structure. The reader will also note that none of the above studies consider random coefficients to account for spatial heterogeneity and temporal dependence effects.

4.1.2 The Empirical Context

There are several approaches to studying and modeling land-use change. Irwin and Geoghegan (2001) and Irwin (2010) provide a good taxonomy of these approaches. In the current study, we derive the empirical discrete choice model based on an economic structural framework for land-use change decisions within a spatially explicit framework. This underlying framework goes beyond mechanistic fitting models for the spatial process of land use change to more closely linking landowner decision behavior to land use patterns. At the same time, we explicitly consider spatial dynamics (caused by interdependence among individual landowners) that lead to the land-use decisions of one landowner affecting that of the landowners of proximally located properties.

To elucidate, consider landowners as being economic agents who make forward-looking inter-temporal land use decisions based on profit-maximizing behavior regarding the conversion of a parcel of land to some other economically viable land use (for example, see Capozza and Li, 1994). The stream of returns from converting a parcel from the current land-use to some other land-use has to be weighed against the costs entailed in the conversion from the current land-use to some other land-use. The premise then is that the land use at any time will correspond to the land use type with the highest present discounted sum of future net returns (stream of returns minus the cost of conversion). Some of the factors affecting the stream of returns and the cost of conversion (and,

[7] Franzese *et al.* (2010) and LeSage and Pace (2009) point out a mistake in the original MCMC method proposed for spatial probit models by LeSage (2000). Essentially, the earlier LeSage (2000) study provided the false perception that Bayesian MCMC was simpler and faster than frequentist methods, because LeSage inadvertently used a univariate truncated normal distribution in translating the latent variables to the observed variables, while a multivariate truncated normal distribution is needed for this purpose. The net result is that the Bayesian MCMC "parallels the computation intensity of the classical RIS strategy" (Franzese *et al.*, 2010).

therefore, the net returns) will be observed (such as road accessibility, distance from flood plain, and the availability and quality of amenities), while others will not.

The net returns may, therefore, be considered to be a latent variable that includes a systematic component and an unobserved component. In addition, spatial interactions are likely to naturally arise because land owners of proximately located spatial units (say, parcels) are likely to be influenced by each other's perceptions of net returns from a certain land-use type investment. These peer influences may be due to strategic or collaborative partnerships between land owners associated with observed variables to the analyst (such as accessibility to city centers and market places) and unobserved variables to the analyst (such as perhaps soil quality and neighborhood attitudes/politics). Such spatial interactions can be captured by relating the latent continuous "net returns" from each land-use type for a parcel (as perceived by the land owner of that parcel) with the corresponding latent "net returns" from surrounding parcels (as perceived by the land owners of those surrounding parcels) using a spatial lag formulation.[8]

But, in addition to the spatial lag-based interaction effect just discussed, it is also likely that there is heterogeneity in the decision-making process of different land owners because of differential responsiveness to various signals relevant to decision-making. For instance, different land owners may perceive the effects of market place proximity on the net returns differently based on their individual experiences, risk-taking behavior, and even vegetation conservation values. This would then translate to a land owner-specific random coefficients formulation for the "net returns", leading to a stationary across-time

[8] Interestingly, many spatial formulations for land-use modeling have considered spatial interactions to be a "nuisance" issue, and have employed a spatial error structure. However, dyadic and related interactions between land owners require the use of a spatial lag structure that allows spillover effects, as also suggested by Carrión-Flores *et al.* (2009). Further, more generally, and as emphasized by McMillen (2010), it is much easier to justify a parametric spatial lag structure when accommodating spatial dependence, while the use of a parametric spatial error structure is "troublesome because it requires the researcher to specify the actual structure of the errors". Beck *et al.* (2006) also find theoretical and conceptual issues with the spatial error model and refer to it as being "odd", because the formulation rests on the "hard to defend" position that "space matters in the error process but not in the substantive portion of the model". As they point out, the implication is that if a new independent variable is added to a spatial error model "so that we move it from the error to the substantive portion of the model", the variable magically ceases to have a spatial impact on neighboring observations. Overall, we submit that land-use models should be developed using the spatial lag formulation or its many variants, and by explicitly linking land owner decision behavior to land use patterns.

correlation in land uses for the same spatial unit. Such land owner-specific random coefficients and resulting temporal correlations of the land-owner's choices across time have been ignored thus far in the literature. Some earlier studies have considered a generic time-stationary random effect (that is, a random coefficient only on the intercept) for each spatial unit in their spatial error formulation, but such a formulation is restrictive relative to the more general random-coefficients spatial lag formulation used here.

In addition to such a general time-stationary random-coefficients effect, there may also be time-varying correlation effects for landowners in their assessment of net returns. Such effects may be due to personality characteristics (such as, say risk averseness or risk acceptance behavior) that fade over time or recent personal experiences.

The implementation of the economic land use change framework discussed above is facilitated by the recent public availability of longitudinal and high resolution spatial land-use data (collected using aerial photography, remote-sensing, and real-estate appraisal information), which enables the modeling of land use at a fine spatial level such as a parcel. In particular, the observed land use data for each spatial unit is in the form of categorical data. Also, the choice of land use is mutually exclusive. Thus, the theoretical "net returns" land use change framework leads naturally to an empirical discrete choice model at a very fine level of spatial resolution (see Bockstael, 1996, Carrión-Flores and Irwin, 2004, Chakir and Parent, 2009, and Carrión-Flores *et al*., 2009). In such a model, the "net returns" concept is replaced by an "instantaneous utility" of each landowner to have a spatial unit in a certain land use type. This utility is a function of exogenous variables and unobserved variables, and the land use observed at a spatial unit corresponds to the one with highest utility. While earlier studies have used such a cross-sectional discrete choice model, no earlier land-use study that we are aware of has considered and applied a discrete choice formulation that simultaneously accommodates the spatial dynamics through a spatial lag structure, spatial heterogeneity through spatial-unit specific random coefficients, time-varying as well as time-stationary unobserved components extracted from multiperiod observations on the same spatial units, as well as

a flexible contemporaneous covariance structure across the utilities of the different land use type alternatives.

4.2 MODELING METHODOLOGY

4.2.1 Model Formulation

Let the instantaneous utility U_{qti} obtained by the landowner of parcel q (q = 1, 2, …, Q) at time t (t = 1, 2, …, T) with land use i (i = 1, 2, …, I) be a function of a (K × 1)-column vector of exogenous attributes x_{qti} . This utility is spatially interdependent across landowners (due to spillover effects based on spatial proximity of parcels) as well as has a temporally interdependent component (due to unobserved factors specific to each landowner). Thus, we write the utility U_{qti} using a spatial lag structure as follows:

$$U_{qti} = \delta \sum_{q'} w_{qq'} U_{q'ti} + \tilde{\alpha}_{qi} + \boldsymbol{\beta}'_q \mathbf{x}_{qti} + \tilde{\varepsilon}_{qti} \quad (1)$$

where $w_{qq'}$ is the usual distance-based spatial weight corresponding to units q and q' (with $w_{qq} = 0$ and $\sum_{q'} w_{qq'} = 1$) for each (and all) q, δ $(0 < \delta < 1)$ is the spatial lag autoregressive parameter, $\tilde{\alpha}_{qi}$ is a normal random-effect term capturing time-stationary preference effects of the landowner of parcel q for land use i, and $\boldsymbol{\beta}_q$ is a parcel-specific (K×1)-vector of coefficients assumed to be a realization from a multivariate normal distribution with mean vector $\mathbf{b}$ and covariance $\tilde{\boldsymbol{\Omega}} = \mathbf{L}\mathbf{L}'$. It is not necessary that all elements of $\boldsymbol{\beta}_q$ be random; that is, the analyst may specify fixed coefficients on some exogenous variables in the model, though it will be convenient in presentation to assume that all elements of $\boldsymbol{\beta}_q$ are random. For later use, we will write $\boldsymbol{\beta}_q = \mathbf{b} + \breve{\boldsymbol{\beta}}_q$, where $\breve{\boldsymbol{\beta}}_q \sim MVN_K(0, \tilde{\boldsymbol{\Omega}})$ (MVN_K represents the multivariate normal distribution of dimension K). Also, for later use, we will write $\tilde{\alpha}_{qi} = \tilde{a}_i + \breve{\alpha}_{qi}$, and let the mean and variance-covariance matrix of the vertically stacked (I×1)-vector of random-effect terms $\tilde{\boldsymbol{\alpha}}_q \left[= \left(\tilde{\alpha}_{q1}, \tilde{\alpha}_{q2}, ..., \tilde{\alpha}_{qI} \right)' \right]$ be $\tilde{\mathbf{A}}$ and $\tilde{\boldsymbol{\Lambda}}$, respectively. $\tilde{\varepsilon}_{qti}$ in Equation (1) is a normal error

term uncorrelated with $\tilde{\boldsymbol{\beta}}_{\boldsymbol{q}}$ and all $\tilde{\alpha}_{qi}$ terms (i = 1, 2, …, I), and also uncorrelated across observation units q. However, the $\tilde{\varepsilon}_{qti}$ terms may have a covariance (dependency) structure across land uses i (due to unobserved factors at time t that simultaneously increase or simultaneously decrease the utility of certain types of land uses) and also a covariance structure across time to recognize time-varying preference effects of the landowner of parcel q. For the time varying effects, it is reasonable to consider that the dependency effects fade over time, and so we consider a first order autoregressive temporal dependency process: $\tilde{\varepsilon}_{qti} = \rho\tilde{\varepsilon}_{q,t-1,i} + \tilde{\eta}_{qti}$, with ρ $(0 < \rho < 1)$ being the temporal autoregressive parameter. The error term $\tilde{\eta}_{qti}$ is temporally uncorrelated, but can be correlated across alternatives - $\tilde{\boldsymbol{\eta}}_{\boldsymbol{qt}} \left[= \left(\tilde{\eta}_{qt1}, \tilde{\eta}_{qt2}, ..., \tilde{\eta}_{qtI}\right)' \right] \sim MVN_I(0, \tilde{\boldsymbol{\Psi}})$. As usual, appropriate scale and level normalization must be imposed on $\tilde{\mathbf{A}}$, $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\boldsymbol{\Psi}}$ for identifiability. Specifically, only utility differentials matter in discrete choice models. Take the utility differentials with respect to the first alternative. Then, only the elements $\alpha_{qi1} = \tilde{\alpha}_{qi} - \tilde{\alpha}_{q1}$ $(i \neq 1)$ and its covariance matrix $\boldsymbol{\Lambda}_1$, and the covariance matrix $\boldsymbol{\Psi}_1$ of $\eta_{qti1} = \tilde{\eta}_{qti} - \tilde{\eta}_{qt1}$ $(i \neq 1)$, are estimable. However, as discussed in Bhat (2011a), the MACML inference approach, like the traditional GHK simulator, takes the difference in utilities against the chosen alternative during estimation. Thus, consider that land use m_{qt} exists at parcel q at time t. This implies that values of $\alpha_{qim_{qt}} = \tilde{\alpha}_{qi} - \tilde{\alpha}_{qm_{qt}}$ $(i \neq m_{qt})$, and the covariance matrices $\boldsymbol{\Lambda}_{m_{qt}}$, and $\boldsymbol{\Psi}_{m_{qt}}$ are desired for parcel q at time t. However, though different random effects differentials and different covariance matrices are used for different parcels and different time periods, all of these must originate in the same values of the undifferenced error term vector $\tilde{\mathbf{A}}$ and covariance matrices $\tilde{\boldsymbol{\Lambda}}$ and $\tilde{\boldsymbol{\Psi}}$. To achieve this consistency, we normalize $\tilde{\alpha}_{q1} = 0 \; \forall \; q$. This implies that $\tilde{a}_1 = 0$. Also, we develop $\boldsymbol{\Lambda}$ from $\boldsymbol{\Lambda}_1$ by adding an additional row on top and an additional column to the left. All elements of this additional row and additional column are filled with values of zeros.

Similarly, we construct $\boldsymbol{\Psi}$ from $\boldsymbol{\Psi}_1$ by adding a row on top and a column to the left. This first row and the first column of the matrix $\tilde{\boldsymbol{\Psi}}$ are also filled with zero values. However, an additional normalization needs to be imposed on $\tilde{\boldsymbol{\Psi}}$ because the scale is also not identified. For this, we normalize the element of $\tilde{\boldsymbol{\Psi}}$ in the second row and second column to the value of one. Note that all these normalizations do not place any restrictions, and a fully general specification is the result. But they are needed for econometric identification.

We now set out notation to write the likelihood function in a compact form. Define the following:

$\mathbf{U}_{qt} = (U_{qt1}, U_{qt2}, ..., U_{qtI})'$, $\tilde{\boldsymbol{\varepsilon}}_{qt} = (\tilde{\varepsilon}_{1t1}, \tilde{\varepsilon}_{qt2}, ..., \tilde{\varepsilon}_{qtI})'$, $\tilde{\boldsymbol{\eta}}_{qt} = (\tilde{\eta}_{qt1}, \tilde{\eta}_{qt2}, ..., \tilde{\eta}_{qTI})'$ ($I \times 1$ vectors),

$\mathbf{U}_q = (\mathbf{U}'_{q1}, \mathbf{U}'_{q2}, ...\mathbf{U}'_{qT})'$, $\tilde{\boldsymbol{\varepsilon}}_q = (\tilde{\boldsymbol{\varepsilon}}'_{q1}, \tilde{\boldsymbol{\varepsilon}}'_{q2}, ...\tilde{\boldsymbol{\varepsilon}}'_{qT})'$, $\tilde{\boldsymbol{\eta}}_q = (\tilde{\boldsymbol{\eta}}'_{q1}, \tilde{\boldsymbol{\eta}}'_{q2}, ...\tilde{\boldsymbol{\eta}}'_{qT})'$ ($TI \times 1$ vectors),

$\mathbf{U} = (\mathbf{U}'_1, \mathbf{U}'_2, ...\mathbf{U}'_Q)'$, $\tilde{\boldsymbol{\varepsilon}} = (\tilde{\boldsymbol{\varepsilon}}'_1, \tilde{\boldsymbol{\varepsilon}}'_2, ...\tilde{\boldsymbol{\varepsilon}}'_Q)'$, $\tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\eta}}'_1, \tilde{\boldsymbol{\eta}}'_2, ...\tilde{\boldsymbol{\eta}}'_Q)'$ ($QTI \times 1$ vectors),

$\breve{\boldsymbol{\alpha}}_q = (\breve{\alpha}_{q1}, \breve{\alpha}_{q2}, ..., \breve{\alpha}_{qI})'$ ($\mathrm{I} \times 1$ vector), $\breve{\boldsymbol{\alpha}} = \left[(1_T \otimes \breve{\alpha}_1)', (1_T \otimes \breve{\alpha}_2)', ...(1_T \otimes \breve{\alpha}_Q)'\right]'$ ($QTI \times 1$ vector), $\mathbf{x}_{qt} = (\mathbf{x}_{qt1}, \mathbf{x}_{qt2}, ..., \mathbf{x}_{qtI})'$ ($I \times K$ matrix), $\mathbf{x}_q = (\mathbf{x}'_{q1}, \mathbf{x}'_{q2}, ..., \mathbf{x}'_{qT})'$ ($TI \times K$ matrix), $\mathbf{x} = (\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}'_Q)'$ ($QTI \times K$ matrix), and $\breve{\boldsymbol{\beta}} = \left(\breve{\boldsymbol{\beta}}'_1, \breve{\boldsymbol{\beta}}'_2, ..., \breve{\boldsymbol{\beta}}'_Q\right)'$ ($QK \times 1$ vector) . Let $\mathbf{IDEN}_E$ be the identity matrix of size E, $\mathbf{1}_E$ be a column vector of size E with all of its elements taking the value of one, and $\mathbf{1}_{EE}$ be a square matrix of size E with all unit elements. Also, define the following matrices:

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} (T \times T \text{ matrix}),$$

$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 & 0 & 0 & 0.\cdots & 0 \\ 0 & \mathbf{x}_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{x}_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & \mathbf{x}_Q \end{bmatrix} \quad (QTI \times QK \text{ matrix}) \text{ d} \tag{2}$$

$\mathbf{S} = \left[\mathbf{IDEN}_{QTI} - \{(\delta \mathbf{W} \otimes \mathbf{IDEN}_T) \otimes \mathbf{IDEN}_I\}\right]^{-1}$ ($QTI \times QTI$ matrix), $\mathbf{W}$ is the $(Q \times Q)$ weight matrix with the weights $w_{qq'}$ as its elements, and $\mathbf{C} = \left[\mathbf{IDEN}_{QTI} - \mathbf{IDEN}_Q \otimes (\rho \mathbf{R} \otimes \mathbf{IDEN}_I)\right]^{-1} = \mathbf{IDEN}_Q \otimes \left(\left[\mathbf{IDEN}_{TI} - (\rho \mathbf{R} \otimes \mathbf{IDEN}_I)\right]^{-1}\right)$ ($QTI \times QTI$ matrix). Then, we can write Equation (1) in matrix notation as:

$$\mathbf{U} = \mathbf{S}\left[\left(\mathbf{1}_{QT} \otimes \tilde{\mathbf{A}}\right) + \mathbf{xb} + \breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \mathbf{C}\tilde{\boldsymbol{\eta}}\right] \tag{3}$$

Let $[.]_e$ indicate the e^{th} element of the column vector [.], and let $d_{qti} = (q-1)TI + (t-1)I + i.$ Equation (3) can be equivalently written as:

$$U_{qti} = \left[\mathbf{S}\left\{\left(\mathbf{1}_{QT} \otimes \tilde{\mathbf{A}}\right) + \mathbf{xb}\right\}\right]_{d_{qti}} + \left[\mathbf{S}\left\{\breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \mathbf{C}\tilde{\boldsymbol{\eta}}\right\}\right]_{d_{qti}} \tag{4}$$

Define $V_{qti} = \left[\mathbf{S}\left\{\left(\mathbf{1}_{QT} \otimes \tilde{\mathbf{A}}\right) + \mathbf{xb}\right\}\right]_{d_{qti}}$ and $\varepsilon_{qti} = \left[\mathbf{S}\left\{\breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \mathbf{C}\tilde{\boldsymbol{\eta}}\right\}\right]_{d_{qti}}$. The landowner of parcel q chooses the land use at time t that provides maximum utility. As earlier, let the land use of parcel q at time t be m_{qt}. In the utility differential form, we may write Equation (4) as:

$$\begin{aligned} &y_{qtim_{qt}} = U_{qti} - U_{qtm_{qt}} = H_{qtim_{qt}} + \xi_{qtim_{qt}}; \\ &H_{qtim_{qt}} = V_{qti} - V_{qtm_{qt}} \quad \textit{and} \quad \xi_{qtim_{qt}} = \varepsilon_{qti} - \varepsilon_{qtm_{qt}};\ i \neq m_{qt} \end{aligned} \tag{5}$$

Then stack the utility differentials $y_{qtim_{qt}} (= U_{qti} - U_{qtm_{qt}}, i \neq m_{qt})$ in the following order: $\mathbf{y_{qt}} = (y_{qt1m_{qt}}, y_{qt2m_{qt}}, ..., y_{qtIm_{qt}})'$, an $(I-1) \times 1$ vector; $\mathbf{y_q} = (\mathbf{y'_{q1}}, \mathbf{y'_{q2}}, ..., \mathbf{y'_{qT}})'$, an $\left[(I-1) \times T\right] \times 1$ vector; and $\mathbf{y} = (\mathbf{y'_1}, \mathbf{y'_2}, ..., \mathbf{y'_Q})'$, an $\left[(I-1) \times T \times Q\right] \times 1$ vector. Correspondingly, let $\mathbf{H_{qt}} = (H_{qt1m_{qt}}, H_{qt2m_{qt}}, ..., H_{qtIm_{qt}})'$, an $(I-1) \times 1$ vector; $\mathbf{H_q} = (\mathbf{H'_{q1}}, \mathbf{H'_{q2}}, ..., \mathbf{H'_{qT}})'$, an $\left[(I-1) \times T\right] \times 1$ vector; and $\mathbf{H} = (\mathbf{H'_1}, \mathbf{H'_2} ..., \mathbf{H'_Q})'$, an $\left[(I-1) \times T \times Q\right] \times 1$ vector. It is easy to see that $\mathbf{y}$ has a mean vector **H.** To determine the

covariance matrix of $\mathbf{y}$, several additional matrix definitions are needed. Define $\mathbf{\Lambda} = \mathbf{IDEN}_Q \otimes \left(\mathbf{1}_{TT} \otimes \tilde{\mathbf{\Lambda}}\right)$ ($QIT \times QIT$ matrix), $\mathbf{\Omega} = \tilde{\mathbf{x}}(\mathbf{I}_Q \otimes \tilde{\mathbf{\Omega}})\tilde{\mathbf{x}}'$ ($QTI \times QTI$ matrix), and $\mathbf{\Psi} = \mathbf{IDEN}_{QT} \otimes \tilde{\mathbf{\Psi}}$ ($QTI \times QTI$ matrix). Let $\tilde{\mathbf{F}} = \mathbf{S}\left[\mathbf{\Lambda} + \mathbf{\Omega} + \mathbf{C}\mathbf{\Psi}\mathbf{C}'\right]\mathbf{S}'$ and define $\mathbf{M}$ as an $[(I-1)\times T \times Q] \times [I \times T \times Q]$ block diagonal matrix, with each block diagonal having $(I-1)$ rows and I columns corresponding to the t^{th} observation time period on parcel q. This $(I-1)\times I$ matrix for parcel q and observation time period t corresponds to an $(I-1)$ identity matrix with an extra column of -1's added as the $m_{qt}{}^{\text{th}}$ column. For instance, consider the case of Q = 2, T = 2, and I = 4. Let parcel 1 be observed to be in land-use 2 in time period 1 and in land-use 1 in time period 2, and let parcel 2 be in land-use 3 in time period 1 and in land-use 4 in time period 2. Then $\mathbf{M}$ takes the form below.

$$\mathbf{M} = \left[\begin{array}{cccc|cccc|cccc|cccc}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1
\end{array}\right] \tag{6}$$

Finally, we obtain the multivariate distribution of the utility differentials $\mathbf{y}: \mathbf{y} \sim \boldsymbol{MVN}(\mathbf{B}, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = \mathbf{M}\tilde{\mathbf{F}}\mathbf{M}'$. Next, let $\boldsymbol{\theta}$ be the collection of parameters to be estimated: $\boldsymbol{\theta} = \left[\mathbf{b}'; \text{Vech}(\tilde{\mathbf{\Omega}}); \tilde{\mathbf{A}}', \text{Vech}(\tilde{\mathbf{\Lambda}}), \text{Vech}(\tilde{\mathbf{\Psi}}), \delta, \rho\right]'$, where $\text{Vech}(\tilde{\mathbf{\Omega}})$ represents the row vector of upper triangle elements of $\tilde{\mathbf{\Omega}}$. Then, the likelihood of the observed sample may be written succinctly as Prob[$\mathbf{y}^* < \mathbf{0}$].

$$L_{ML}(\boldsymbol{\theta}) = \text{Prob}[\mathbf{y}^* < 0] = F_{Q \times T \times (I-1)}(-\mathbf{B}, \mathbf{\Sigma}) \tag{7}$$

where $F_{Q\times T\times(I-1)}$ is the multivariate cumulative normal distribution of $Q\times T\times(I-1)$ dimensions. Despite advances in simulation techniques and computational power, the evaluation of such a high dimensional integral is literally infeasible using traditional frequentist and Bayesian simulation techniques. For instance, in frequentist methods, where estimation is typically undertaken using pseudo-Monte Carlo or quasi-Monte Carlo simulation approaches (combined with a quasi-Newton optimization routine in a maximum simulated likelihood (MSL) inference), the computational cost to ensure good asymptotic estimator properties can be prohibitive as the number of dimensions of integration increases (see Bhat *et al*., 2010b for a detailed discussion of frequentist simulation procedures and problems under high integration dimensionality). Similar problems arise in Bayesian Markov Chain Monte Carlo (MCMC) simulation approaches, which remain cumbersome, require extensive simulation, are time consuming, and pose convergence assessment problems as the number of dimensions increases (see Müller and Czado, 2005, Ver Hoef and Jansen, 2007, and Franzese *et al*., 2010 for discussions). The surrogate CML function for the likelihood in equation (7) may be written as:

$$L_{CML}(\boldsymbol{\theta})=\prod_{q=1}^{Q}\prod_{q'=q}^{Q}\prod_{t=1}^{T}\prod_{t'=t}^{T}\text{Prob}\left(C_{qt}=m_{qt},C_{q't'}=m_{q't'}\right)\text{with}\,q\neq q'\,\text{when}\;\;t=t', \tag{8}$$

where C_{qt} is an index for the land use in which parcel *q* is at time *t*. Each of these pairwise probabilities is of $(I-1)\times 2$ dimensions, which may be computed easily using the MVNCD approximation method embedded in the MACML method.[9]

The pairwise marginal likelihood function of Equation (8) comprises $QT(QT-1)/2$ pairs of pairwise probability computations. But, in a spatial-temporal case

[9] It should be noted that while the MVNCD approximation of Equation (8) provides a relatively simple objective function to be maximized with respect to the parameters, the resulting function can theoretically have multiple maxima (note, however, that this is also true of the likelihood function of pretty much every other multinomial discrete choice model except the multinomial logit model). There is no way out of this "multiple maxima" situation, other than for the analyst to test various different starting points and see whether the parameters converge to the same point. We undertook such an analysis with some of the data sets generated as part of testing whether the MACML procedure is able to recover parameters (see next section), and found that the parameters always converged to the same point. Of course, this does not mean that there are no multiple optima, because it is impossible to test the infinite number of possible starting parameter spaces; but the testing does suggest reasonable stability of the maximization procedure.

where spatial dependency drops quickly with inter-observation distance, the pairs formed from the closest observations provide much more information than pairs that are very far away. In fact, as demonstrated by Varin and Vidoni (2009), Bhat *et al*. (2010a), and Varin and Czado (2008) in different empirical contexts, retaining all pairs may reduce estimator efficiency. We examine this issue by creating different distance bands (including the band that includes all pairings) and, for each specific distance band, considering only those unordered pairings in the CML function that are within the distance band. Then, we develop the asymptotic variance matrix $V_{CML}(\hat{\boldsymbol{\theta}})$ for each distance band and select the threshold distance value (say $\tilde{d}_{thresh}$) that minimizes the total variance across all parameters as given by $tr[V_{CML}(\hat{\boldsymbol{\theta}})]$ (*i.e.*, the trace of the matrix $[V_{CML}(\hat{\boldsymbol{\theta}})]$.[10]

The CML estimator of $\boldsymbol{\theta}$ is consistent and asymptotically normal distributed with asymptotic mean $\boldsymbol{\theta}$ and covariance matrix given by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005):

$$V_{CML}(\hat{\boldsymbol{\theta}}) = [G(\boldsymbol{\theta})]^{-1} = [H(\boldsymbol{\theta})]^{-1} J(\boldsymbol{\theta}) [H(\boldsymbol{\theta})]^{-1}, \text{ where} \quad (9)$$

The "bread" matrix $H(\boldsymbol{\theta})$ of Equation (9) can be estimated in a straightforward manner using the Hessian of the negative of the MACML likelihood function, evaluated at the MACML estimate $\hat{\boldsymbol{\theta}}$. On the other hand, the "vegetable" matrix $J(\boldsymbol{\theta})$ is not that straightforward to estimate. But the decaying nature of the distance weight matrix can be used to create pseudo-independent subsamples of the data using the windows sampling method proposed by Heagerty and Lumley (2000). Based on this windows sampling method, Bhat (2011a) suggests overlaying the spatial region under consideration with a square grid providing a total of *D* internal and external nodes. Then, select the observational unit closest to each of the *D* grid nodes to obtain *D* observational units

[10] We do not test different time period bands like we do distance bands. This is because, unlike the spatial dependency pattern, the temporal dependency includes a time-invariant component that does not fade over time. Thus, for each observation unit, all time periods need to be considered, unless the time-stationary dependency is negligible, which we generally do not believe will be the case.

from the original Q observational units ($\tilde{d} = 1, 2, 3, \ldots, D$). Let $\tilde{\mathbf{C}}$ be a $Q \times D$ matrix with its $\tilde{d}^{th}$ column filled with a $Q \times 1$ vector of 0s and 1s, with a zero value in the q'^{th} row ($q' = 1, 2, \ldots, Q$) if the observational unit q' is not within the specified threshold distance $\tilde{d}_{thresh}$ of unit $\tilde{d}$, and a one otherwise (by construction, $\tilde{\mathbf{C}}_{q'\tilde{d}} = 1$ if $q' = \tilde{d}$). Also, let $\breve{\mathbf{C}} = 1_T \otimes \tilde{\mathbf{C}}$. Then, the columns of $\breve{\mathbf{C}}$ provide pseudo-independent sets of observational units.[11] Let the score matrix corresponding to the pairings in column $\tilde{d}$ of matrix $\breve{\mathbf{C}}$ be $S_{CML,d}(\boldsymbol{\theta})$. Also, Let $N_{\tilde{d}}$ be the sum of the $\tilde{d}^{th}$ column of $\breve{\mathbf{C}}$, and let $\tilde{W}$ be the total number of pairings used in the CML function of Equation (8) (after considering the distance threshold $\tilde{d}_{thresh}$). Then, the **J** matrix maybe empirically estimated as:

$$\mathbf{J} = \frac{\tilde{W}}{D}\left[\sum_{d=1}^{D}\left[\frac{1}{N_d}\left([S_{CML,d}(\boldsymbol{\theta})][S_{CML,d}(\boldsymbol{\theta})]'\right)_{\hat{\theta}}\right]\right]. \tag{10}$$

One additional issue regarding estimation. The analyst needs to ensure the positive definiteness of the three covariance matrices, $\tilde{\boldsymbol{\Omega}}$, $\tilde{\boldsymbol{\Lambda}}$, and $\tilde{\boldsymbol{\Psi}}$. Once this is ensured, and as long as $0 < \rho < 1$ and $0 < \delta < 1$, $\boldsymbol{\Sigma}$ will be positive definite. In the estimation, the positive-definiteness of each of the $\tilde{\boldsymbol{\Omega}}$, $\tilde{\boldsymbol{\Lambda}}$, and $\tilde{\boldsymbol{\Psi}}$ matrices is guaranteed by writing the logarithm of the pairwise-likelihood in terms of the Cholesky-decomposed elements of these matrices, and maximizing with respect to these elements of the Cholesky factor. Essentially, this procedure entails passing the Cholesky elements as parameters to the optimization routine, constructing the covariance matrix internal to the optimization routine, then computing $\boldsymbol{\Sigma}$, and finally picking off the appropriate elements of the matrix for the pairwise likelihood components. To ensure the constraints on the autoregressive

[11] As indicated by Bhat (2011a), there needs to be a balance here between the number of sets of pairings *D* and the proximity of points. The smaller the value of *D*, the less proximal are the sets of observation units and more likely that the sets of observational pairings will be independent. However, at the same time, the value of *D* needs to be reasonable to obtain a good empirical estimate of *J*, since this empirical estimate is based on averaging the cross-product of the score functions (computed at the convergent parameter values) across the *D* sets of observations.

terms ρ and δ, we parameterize these terms as $\rho = 1/[1+\exp(\tilde{\rho})]$ and $\delta = 1/[1+\exp(\tilde{\delta})]$, respectively. Once estimated, the $\tilde{\rho}$ and $\tilde{\delta}$ estimates can be translated back to estimates of ρ and δ.

4.2.2 Simulation Study

We undertake a simple simulation exercise to examine the ability of the MACML estimation approach to recover the parameters. A four-alternative choice situation (I = 4) with four time periods (T = 4) is considered for the simulation exercise (this scenario matches with the dimensions of the empirical study here). A total of Q = 200 observation units are assumed (the observation units correspond to parcels in this case). Three independent variables are used in the utility equation and each of them are generated from a standard univariate normal distribution (these are the elements of the x_{qti} vector). A random coefficient (across observation units) is assumed on the first variable, while fixed coefficients are assumed on the other two variables. Observation-specific random effects are also introduced, with the mean effect for the first alternative normalized to zero (equivalent to setting $\tilde{a}_1 = 0$). A diagonal specification is considered for the covariance matrix of $\tilde{\mathbf{\Lambda}}$, with the assumption that there is no random effect for the first alternative (note, however, that, as discussed in Section 4.2.1, a more general covariance specification is estimable for $\tilde{\mathbf{\Lambda}}$). In particular, using the normalization procedure in Section 4.2.1., $\tilde{\mathbf{\Lambda}}$ is specified as:

$$\tilde{\mathbf{\Lambda}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The covariance matrix $\tilde{\mathbf{\Psi}}$ for the error term vector $\tilde{\boldsymbol{\eta}}_{qt}$ is also diagonal, but now with the fixed value of 0.5 along the diagonal for each alternative. Such a matrix is again a restrictive case of the more general $\tilde{\mathbf{\Psi}}$ covariance matrix discussed in Section 4.2.1. Note that such a structure simplifies the simulation, since the elements of $\tilde{\mathbf{\Psi}}$ are not estimated.

The reason for such restrictions on the $\tilde{\mathbf{\Lambda}}$ and $\tilde{\mathbf{\Psi}}$ matrices in the simulation design is to restrict the number of parameters to be estimated (given the moderate size sample used in the experiments) and to focus on the spatial and temporal dependency patterns.

To generate the spatial lag dependency, the 200 observation points are located on a rectangular grid of size 3,800 meters (2.375 miles) by 1800 meters (1.125 miles). Each observation point is 200 meters away from its closest neighbor. The spatial weight matrix W (of size 200×200) is created using the inverse of the square of distance on the coordinate plane between observational units. Finally, the first-order AR(1) temporal dependency parameters ρ is specified to be 0.6, and the spatial lag parameter δ is specified to be 0.5.

In total, the simulation design includes 13 parameters: three mean coefficients on the exogenous variables (corresponding to the **b** coefficient vector), one random coefficient element (corresponding to the Cholesky matrix in $\tilde{\mathbf{\Omega}} = \mathbf{L}\mathbf{L}'$, which in our simulation design is the standard deviation of $\breve{\beta}_1$), three mean coefficients for the random effects (corresponding to the $\tilde{\mathbf{A}}$ vector), three parameters for the random effects (corresponding to the Cholesky decomposition of the $\tilde{\mathbf{\Lambda}}$ matrix, which in our simulation design are the standard deviations of $\breve{\alpha}_2$, $\breve{\alpha}_3$, *and* $\breve{\alpha}_3$), and the ρ and δ parameters. The simulation experiments entail assuming underlying "true" values for these parameters and generating data sets for estimation. Specifically, using the pre-specified parameters, the utility vector **U** of Equation (3) is generated by drawing realizations of $\breve{\boldsymbol{\beta}}$, $\breve{\boldsymbol{\alpha}}$, and $\tilde{\boldsymbol{\eta}}$ from their underlying distributions. Then, for each observation unit and choice occasion, the alternative with the highest utility is designated as the chosen alternative. This variable constitutes the discrete dependent variable. The above procedure is repeated 50 times with different realizations of the $\breve{\boldsymbol{\beta}}$, $\breve{\boldsymbol{\alpha}}$, and $\tilde{\boldsymbol{\eta}}$ to generate 50 different data sets. For each dataset, the 13 model parameters are estimated using the MACML method, considering all pairs of observations in the CML function.

The results from the estimations are translated to measures of performance by comparing the estimated parameters with the "true" parameter values. To evaluate the

ability of the MACML procedure to recover the parameters accurately, we compute an absolute percentage bias (APB) measure for each parameter, which is the deviation of the mean estimate for the parameter. Table 4.1 presents the results. The MACML inference approach does quite well in recovering parameters for all distance bands, with the mean APB value being around 6% (see last row of the APB column). More extensive theoretical and simulation studies are needed to better understand and characterize the ability of the MACML estimator to recover parameters under alternative spatial-temporal dependency scenarios, and also to investigate estimator efficiency considerations. But, the simple simulation exercise undertaken here suggests that the MACML method is able to recover the true parameters remarkably well for the spatial lag unordered response model with temporal autocorrelation.

Table 4.1. MACML Estimation Results of 50 Simulated Datasets with 200 Individuals and 4 Time Periods

Parameter		**True**	**MACML estimate**		
Notation in Formulation	**Components**		**Mean est.**	**Abs. Bias**	**Abs. %age bias (APB)**
B	b1	0.500	0.521	0.021	4.2%
	b2	0.700	0.710	0.010	1.5%
	b3	-0.600	-0.627	0.027	4.5%
Chol($\tilde{\boldsymbol{\Omega}}$)	$SD-\breve{\beta}_1$	1.000	1.033	0.033	3.3%
$\tilde{\mathbf{A}}$	a2	0.300	0.334	0.034	11.3%
	a3	-0.400	-0.385	0.015	3.7%
	a4	0.500	0.569	0.069	13.8%
Chol($\tilde{\boldsymbol{\Lambda}}$)	$SD-\breve{\alpha}_2$	1.000	1.047	0.047	4.7%
	$SD-\breve{\alpha}_3$	1.000	1.009	0.009	0.9%
	$SD-\breve{\alpha}_4$	1.000	1.119	0.119	11.9%
ρ	rho	0.600	0.576	0.024	4.0%
δ	del	0.500	0.465	0.035	6.9%
Mean value across parameters				**0.037**	**5.9%**

4.3 APPLICATION

4.3.1 The Data and the Context

The data used in this study is from the City of Austin, Texas. Parcel level land use inventory data for the years 1995, 2000, 2003 and 2006 are used. This data is available in the Environmental Systems Research Institute's (ESRI's) shape file format for all the four years for a 2 mile extraterritorial jurisdiction (ETJ) of the City of Austin, covering a total area of 1795 sq km. (693 sq mile). The land use type for each parcel is available at a fine level of detail; however, for the current study, they are aggregated into four mutually exclusive land use categories. These are (1) residential (including single family, duplexes, three/four-plexes, apartments, condominiums, mobile homes, group quarters, and retirement housing), (2) commercial (including commercial, office, hospitals, government services, educational services, cultural services, and parking), (3) industrial (including manufacturing, warehousing, resource extraction (mining), landfills, and miscellaneous industrial), and (4) undeveloped (including open and undeveloped spaces, preserves, parks, golf courses, and agricultural open spaces).

An area measuring 23.5 sq km (9.06 sq miles) in the suburbs of Austin city is selected for the analysis. The interstate highway, IH-35, divides this analysis area into an eastern section with two-thirds of the total area and a western section with the remaining one-third of the area. A part of the eastern section falls within the City of Pflugerville, a suburban Austin city. Mopac (Loop 1), another major expressway in Austin, also runs in the North-South direction about half a mile west of the western boundary of the analysis area. Apart from IH-35 and Mopac, three minor arterials and a major arterial pass through the analysis area. All the explanatory variables were created from the GIS data obtained from the City of Austin, except the flood plains data, which was obtained from the Capital Area Council of Governments (CAPCOG).

For the econometric analysis, the area is divided into 400 square cells each of size 242m×242m. The land use in the parcel at the centroid of each cell is designated as the land use for that grid cell. If the centroidal point falls well within the right-of-way of an arterial roadway or other roadways with high land-use access functionality, the

corresponding grid cell is assigned the predominant land use of the adjacent area. However, if the centroidal point of a grid cell falls within the right-of-way of IH-35, which primarily serves the functionality of through movement, we removed the corresponding grid cell from analysis (a total of five grid cells were accordingly removed, leaving a sample of 395 grid cells observed at each of four time points).

In the rest of this study, and for ease in presentation, we will use the terms "grid cell" and "parcel" interchangeably, though the analysis is technically being conducted at the grid cell level. The explanatory variables for each parcel considered in the model include road access measures (distance to IH-35, distance to Mopac, distance to the nearest non-freeway roadway, and interactions of these variables), location relative to the flood plains, an interaction term of proximity to road access with proximity to the flood plain (distance to nearest road divided by distance to the nearest flood plain), being situated in Pflugerville city, and proximity to schools.[12] To construct distances (all measured in kilometers) from each parcel to the roadways, a road network data in polyline format (obtained from the City of Austin) was overlaid on the analysis area, and the Euclidean distance from the parcel to roadways was calculated. To construct distances from each parcel to the nearest flood plain, the flood plain data in polygon format (obtained from the Capital Area Council of Governments) was overlaid onto the analysis area, and Euclidean distances were computed from each parcel centroid to the nearest floodplain polygon. School data was available as point data, and this was overlaid on the analysis area to obtain the distance from a parcel to the nearest school.

Among the exogenous variables considered, we expect that land-owners of parcels in close proximity to highways will most likely invest their parcels in commercial and industrial land-uses. On the other hand, one can expect parcels located far from highways and roadways to remain undeveloped, as land-owners are not likely to see much net returns in developing these parcels. Similarly, we can expect parcels in close proximity to flood plains not to be built up. In addition, we consider an interaction effect

[12]A floodplain is an area susceptible to flooding. Such areas in the United States are identified by the Federal Emergency Management Agency (FEMA) in its Flood Insurance Rate Maps, which show spatial regions likely to be affected by a 100-year flood (1% chance of a flood of this magnitude during the year).

of distance to the nearest roadway divided by distance to the nearest flood plain. This captures the potential "push-pull" non-linear positive effect (on the propensity of a parcel being undeveloped) of being afar from roadways *and* being proximal to a flood plain. However, the land-owner of a parcel that is distant from roadways may see some "net returns" potential in developing the parcel if the parcel is also far away from the flood plains. Similarly, the land-owner of a parcel that is close to a flood plain may still invest the parcel in some kind of development if the parcel is close to roadways. All of these effects are captured by introducing the "distance to nearest roadway divided by distance to the nearest flood plain" variable. The Pflugerville city dummy variable is introduced to capture the effects of a differential development/tax incentive structure in Pflugerville relative to the remainder of the analysis region. Finally, the proximity to schools is likely to be an incentive to develop the parcel for residential land-use.

Figure 4.1 shows the analysis area along with the roadways in the region, the boundary of Pflugerville, the locations of the flood plains and the land-use type of each grid point for the year 1995 (see the legend for land-use type at the bottom right of the figure). Figure 4.2 is the corresponding figure for the year 2000. Several observations may be made just from a visual scan of the figures. First, there is a clustering of parcels in industrial and commercial land-uses immediately adjacent to IH-35. Second, there are more parcels in an undeveloped state as one goes eastwards, away from Mopac. Third, parcels close to the floodplains indeed are more likely to be in an undeveloped land-use state. Fourth, the share of parcels within Pflugerville city in commercial land-use appears higher than in other areas of the analysis region. Fifth, while residences are not necessarily closely clustered around each school, there is a tendency to have quite a few residential parcels within a reasonable range of schools (this visual scan suggests the need to test distance bands from parcels to schools rather than a simple continuous representation of distance from school). Sixth, there is clear evidence of parcels with the same land-use in close proximity, reinforcing the notion of spatial dynamics at play. This effect is particularly obvious when looking at each of the eastern and western sections of the area (as delineated by IH-35) individually.

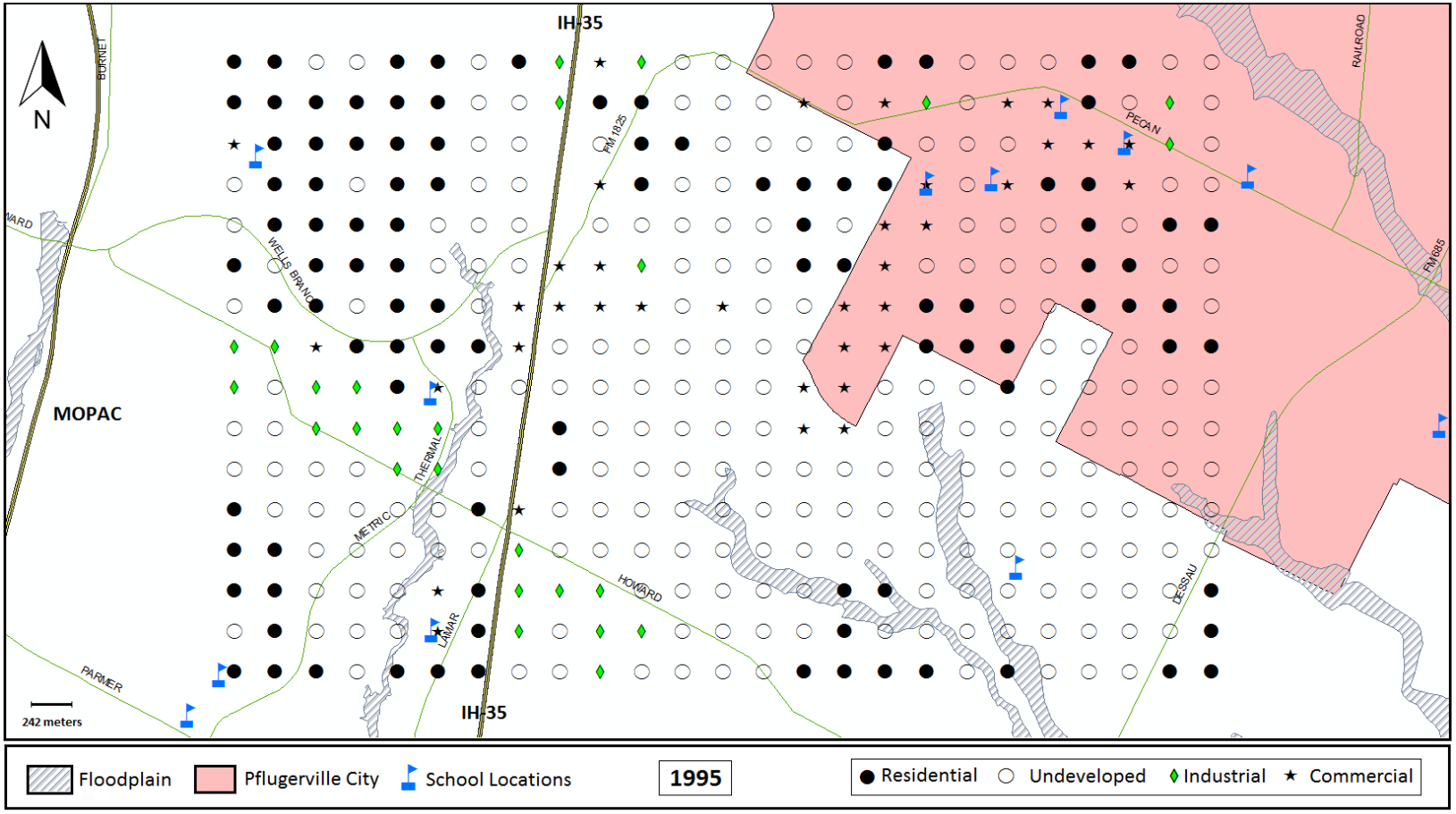

Figure 4.1. The Analysis Area for the year 1995

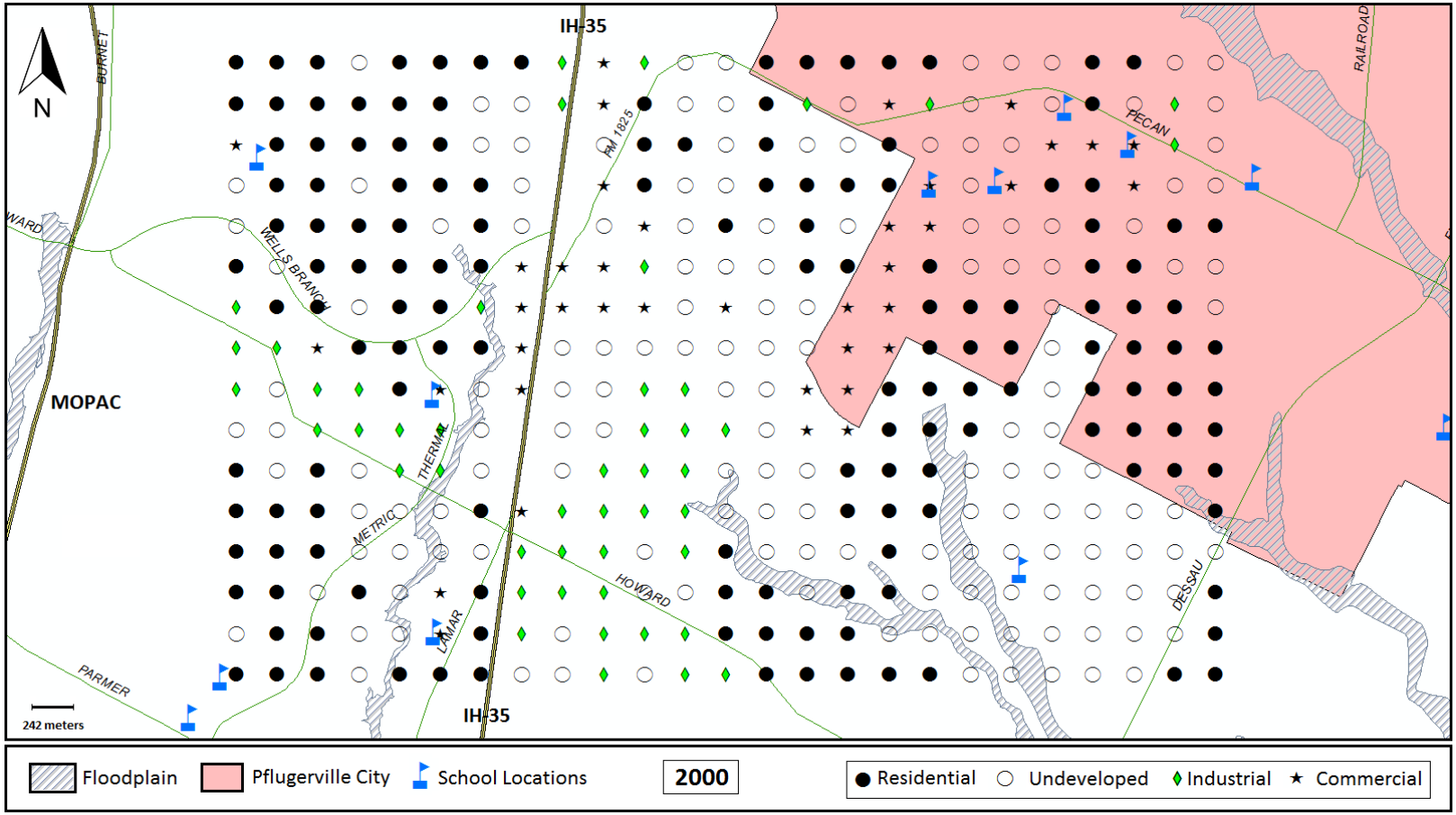

Figure 4.2. The Analysis Area for the year 2000

Seventh, one can see how the clustering effect of similar land-uses manifests itself in the change from 1995 to 2000. Specifically, it can be clearly observed from Figures 4.1 and 4.2 that many parcels in an undeveloped state in 1995 are in a developed state in 2000. Most of these conversions are to residential land-use, though there also is a clear surge in industrial land-use in 2000. As can be noted, there is a distinct clustering pattern in parcels that change from an undeveloped state to each of the residential and industrial land-use types.

Table 4.2 shows the percentage shares of parcels in each of the four land use types at each of the four years of analysis. A high share of the parcels is either in residential or undeveloped land-uses, with the commercial and industrial land-uses representing about 20% of the total share. Another observation from this table, also visible from Figures 4.1 and 4.2, is the boom in residential land development that occurred between 1995 and 2000 (and the reduction in the share of undeveloped land during the same period). This boom is consistent with ground reality in the Austin region (Glaeser *et al.*, 2006). Historically speaking, Austin, like other cities in Texas, has had relatively weak land use zoning policies. Thus, the economic prosperity of the late 90s (and into the first year of the new millennium) led to substantial and relatively uncontrolled development in the Austin area, resulting in the emergence of several low density residential enclaves at the fringes of the main city (such as the area considered in this study). Of course, this growth tapered off and came to a literal standstill after 2001 (see also Table 4.2), attributable to the economic recession that began around March 2001 rather than to any land-use regulations.[13]

[13] To be sure, Austin has had comprehensive development plans since 1928, including the Austin Tomorrow Comprehensive Plan (ATCP) adopted by the City Council in the late 1970s. The ATCP was not acted upon due to lack of consensus and shifts in the Austin City Council make-up over the years, furthered by the non-involvement of the Austin development community. The ATCP was resurrected in 2008 with interim updates. In the meantime, the Austin City also established an explicit and streamlined land use planning process with significant public participation to develop a future land use map (FLUM), which provides the framework for zoning regulations (see City of Austin, 2008). But these issues are not relevant for the period of analysis in the current study.

Table 4.2. Percentage of Land by Land Use Types

Land use Type	1995	2000	2003	2006
Residential	25.80%	38.70%	39.70%	39.50%
Undeveloped/Open Area	58.20%	39.50%	42.50%	39.70%
Commercial	9.40%	9.90%	9.90%	13.40%
Industrial	6.60%	11.90%	7.80%	7.30%

4.3.2 Variable Specification and Spatial Weight Matrix Formulation

Many different variable specifications, functional forms, and variable interactions were considered to determine the final model specification. The roadway access variables (distance from IH-35, Mopac, and other arterials) as well as the distance to the closest flood plain polygon were considered both in linear and non-linear forms (such as the logarithm of distance, the square of distance, and spline variables that allow piece-wise linear effects of distance on the utilities). In addition, we also considered dummy variables for different ranges of distance for these variables (for instance, parcel is within 200 meters of IH-35, parcel is within 300 meters of IH-35, *etc.*). The Pflugerville city location dummy variable was introduced as a switch variable taking the value of '1' for parcels within the City of Pflugerville and '0' otherwise. The proximity to school effect was considered similar to the other continuous variables, and included alternative functional forms of distance from the nearest school as well as dummy variables for different ranges of distance from school (such as parcel is located within 300 meters of a school and within 1 kilometer of a school). In addition to the variables just discussed, we also included a "1995 dummy variable" to capture the rather substantial temporal shifts in shares among the land-use categories between this first year and the subsequent years (see Table 4.2). Further, various interactions of the continuous and the categorical variables were also considered whenever adequate observations were available to test such interaction effects. The final model specification was obtained after extensive explorations and testing, and based on statistical fit, intuitiveness, parsimony considerations, and the preliminary insights offered by the visual scan of Figure 4.1 (as discussed in the previous Section). Specifically, in terms of statistical fit, we used the

adjusted composite likelihood ratio test (AD*CLRT*) statistic (see Pace *et al*., 2011 and Bhat, 2011a) to compare nested models and the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (2005) to test non-nested models.

Table 4.3 provides the descriptive statistics of the independent variables in the final model specification. We also examined alternative specifications for the construction of the spatial weights, including inverse distance and the inverse of the square of distance, the inverse of exponential distance, a simple contiguity indicator, and a contiguity weight but based on shared boundary length rather than a simple indicator. Further, based on the insights from the visual scan of Figure 4.1, we decided to test two spatial variants for the weight specification.

Table 4.3. Descriptive Statistics of the Independent Variables used in the Model

Variable	Min	Max	Mean	Std. Deviation
Distance to IH-35 less than 350 meters	0.0000	1.0000	0.1038	0.3051
Distance to IH-35	0.0440	4.2664	1.7379	1.1556
Distance to nearest non-freeway road (other than IH-35 and Mopac)	0.0004	1.7308	0.5397	0.4254
Distance to IH-35 * Distance to nearest non-freeway road	0.0600	4.3456	1.1479	1.0903
Distance to Mopac Freeway	0.8150	7.154	3.9816	1.7321
Distance to the nearest Roadway	0.0004	1.7308	0.5191	0.4262
Distance to the nearest Flood Plain	0.0012	1.6492	0.6257	0.4065
Distance to the nearest Road / Distance to the nearest flood plain	0.0006	11.5384	1.4367	2.0227
Parcel lies within Pflugerville City	0.0000	1.0000	0.2278	0.4196
Within one kilometer of a school	0.0000	1.0000	0.6127	0.4873

The first was to develop the weight matrix between any two parcels over the entire analysis region. The second was to assume no spatial dependency between parcels on the western and eastern sections of the analysis area (as determined by IH-35), but

assuming dependence between parcels within each of the two sections. That is, if two parcels are located on the same side of IH-35, then the spatial weight for the pair is non-zero based on the weight matrix; otherwise, the spatial weight for the pair is assigned a value of zero. At the end of this extensive testing, which was undertaken using all pairwise interactions in the CML function, the best weight specification involved the inverse of the square of distance specification with spatial dependency confined to parcels within each of the western and eastern sections of the analysis area (with no dependence between parcels lying on opposite sides of I-35). This selection from among the many non-nested weight specifications was undertaken using the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (2005), which takes the following form:

$$CLIC = \log L_{CML}(\hat{\boldsymbol{\theta}}) - tr\left[\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})^{-1}\right] \tag{11}$$

where $\hat{\boldsymbol{\theta}}$ represents the estimated model parameter vector, and $\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{H}}(\hat{\boldsymbol{\theta}})$ are the estimated "vegetable" and "bread" matrices as discussed in Equations (9) and (10), respectively. The model that provides a higher value of CLIC is preferred. For instance, Table 4.4 provides the values of the log-composite likelihood at convergence $\log L_{CML}(\hat{\boldsymbol{\theta}})$, the trace value in the CLIC statistic ($tr(\mathbf{J}(\hat{\boldsymbol{\theta}})\mathbf{H}(\hat{\boldsymbol{\theta}})^{-1})$), and the CLIC statistic value for the model that constructs the weight matrix over the entire region (Full region-based weight matrix model) and the model that constructs the weight matrix over each of the eastern and western sections of the analysis region (Partitioned region-based weight matrix model), with the preferred inverse of the square of distance as the basis for the weight matrix. As can be observed from the CLIC statistic column, the partitioned region-based spatial weight matrix model is superior in representing spatial effects in the current empirical context, indicating the lack of dyadic interactions between land-owners of parcels on either side of IH-35. This result emphasizes the social separation that can be caused by a physical barrier such as a freeway.[14] Finally, using the preferred combination

[14] One of the reviewers encouraged us to develop a more rigorous confidence level-based procedure to test this social barrier hypothesis, and suggested a bootstrapping of the CLIC statistic. To do so, we generate

of the variable specification, and the partitioned weight matrix with the inverse of the square of distance as the separation measure, we undertook an efficiency analysis to determine the optimal distance band for including pairwise interactions in the CML function, based on minimizing the trace of the variance-covariance matrix given by $tr[\mathbf{V}_{CML}(\hat{\boldsymbol{\theta}})]$ (see Section 4.2.2). The $tr[\mathbf{V}_{CML}(\hat{\boldsymbol{\theta}})]$ value was the lowest for a distance band of 400 meters (other distance bands considered included 800, 1200 and 7000 meters, the last one representing the case of including all pairs of parcel-year observations in the CML function). Thus, all subsequent results for models including spatial dependency are based on the 400 meters distance band.

Table 4.4. Model Selection

Statistic	**Full region-based weight matrix model**	**Partitioned region-based weight matrix model**
Log-composite likelihood at convergence	-2568369	-2567958
Trace value	171	136
CLIC statistic	-2568540	-2568094

The next section discusses the results of the following two models in more detail: (1) the multinomial probit model with no temporal and spatial dependencies or the MNP model (in the notation of Section 4.2.1, this model imposes the restrictions that $\tilde{\boldsymbol{\Omega}} = \mathbf{L}\mathbf{L}'$ is a $K \times K$-matrix of zero values, $\tilde{\boldsymbol{\Lambda}}$ is an $I \times I$-matrix of zeros, $\rho = 0$, and $\delta = 0$), and (2) the multinomial probit model with temporal and spatial dependencies (MNPTS). In both of these models, we could not reject the null hypothesis that, after accommodating the exogenous variables, the covariance matrix $\tilde{\boldsymbol{\Psi}}$ had the structure below:

---

100 data sets using the estimated values for the partitioned region-based weight matrix model (PRWM) model. Then, for each data set, we estimate the full region-based weight matrix model (FRWM) and the PRWM model, and subsequently obtain the corresponding CLIC statistics (say, CLIC-FRWM and CLIC-PRWM). In 75% of the bootstrap-generated data sets, we obtained CLIC-PRWM>CLIC-FRWM, providing confidence that the "social barrier" finding is not simply an artifact of sampling. We would like to thank the referee for suggesting that we pursue such an effort.

$$\tilde{\boldsymbol{\Psi}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0.5 & 0.5 \\ 0 & 0.5 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{bmatrix}, \tag{12}$$

which is equivalent to the specification that the intrinsic utility preferences are independent and identically distributed across the four alternatives (with the scale normalized to 0.5). However, note that the MNPTS model does incorporate both dependence and heteroscedasticity across the overall utilities of the alternatives because of the random coefficients on the exogenous variables. Finally, in the MNPTS model, we could not reject the hypothesis that the covariance matrix $\tilde{\boldsymbol{\Lambda}}$ had the following form (the utilities are arranged in the following order of land-use type: residential, commercial, industrial, and undeveloped):

$$\tilde{\boldsymbol{\Lambda}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \Lambda_{com} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Lambda_{und} \end{bmatrix}, \tag{13}$$

The covariance matrix above indicates that there are no time-stationary random effects in the utilities for the residential and industrial land-uses. More intuitively speaking, land-owners are likely to have intrinsic (unobserved and randomly distributed) time-invariant utility "biases" (or preferences) for commercial and undeveloped land-use types, but not for residential and industrial land-use types. This also implies that the utilities for commercial and undeveloped land-use types are correlated across time due to time-invariant land-owner preferences.

4.3.3 Model Estimation Results

The results of the MNP and the MNPTS models are presented in Table 4.5. We first discuss the effects of variables on the utilities of alternatives (Section 4.3.3.1), next the temporal and spatial effects (Section 4.3.3.2), then the model fit comparisons (Section 4.3.3.3), and finally the variable magnitude effects (Section 4.3.3.4). A '-' entry in a cell

of Table 4.5 indicates that the corresponding "row" variable did not have a statistically significant effect on the utility of the corresponding "column" land-use category.

4.3.3.1 Variable Effects on Utility of Alternatives

The estimated coefficients of the two models in Table 4.5 are not directly comparable, since the scales of the error terms in the utilities are different. But the mean coefficient estimates are the same in sign in both models. All the results are consistent with the hypotheses in Section 4.4.3.1. The constant terms do not have any substantive interpretations, and simply represent adjustments in the utilities of alternatives after accommodating the other variables in the model. The presence of standard deviations on the constants for the commercial and undeveloped land-uses (in the MNPTS model) indicates time-invariant preference heterogeneity across landowners in the utilities for these land-uses, as discussed earlier. Parcels located proximal to IH-35 are more likely to be invested in commercial and industrial land-uses, though the functional form of proximity to IH-35 in the utilities of these two land-uses takes different forms. For commercial land-use, the proximity to IH-35 enters as a distance band of 350 meters from IH-35, which is consistent with the clustering of commercial parcels close to IH-35 in Figure 4.1. However, for industrial land-use, the linear form of distance to IH-35 (interacted with distance to nearest non-freeway road) enters the utility function, again consistent with the relative scatter of industrial parcels around IH-35. More generally, industrial facilities (and therefore their land-owners) gain from proximity to freeways. At the same time, zoning setback guidelines can preclude owners of parcels that are immediately adjacent to freeways from investing their land in industrial use (which is why the distance band specification did not come out statistically significant for the industrial land-use alternative). Also, land-owners of parcels close to other major roads can benefit from placing their land in industrial use because of improved transportation accessibility. These behaviors are captured by the negative coefficient for the industry land use category on the interaction variable of the distance to IH-35 and the distance to the nearest non-freeway road.

Table 4.5. Estimation Results (t-statistics in parenthesis)

Variables	Standard multinomial probit (MNP) model				MNP Spatial lag model with temporal Panel and spatial effects (MNPTS) model autocorrelation			
	Residential	**Commercial**	**Industrial**	**Undeveloped**	**Residential**	**Commercial**	**Industrial**	**Undeveloped**
Constant	-	-0.864 (-15.47)	-0.133 (-2.31)	-0.269 (-3.38)	-	-1.869 (-3.90)	0.471 (5.64)	-0.597 (-2.20)
Standard deviation	-	-	-	-	-	2.353 (1.25)	-	2.403 (1.46)
Distance to IH-35 less than 350 meters	-	1.090 (12.97)	-	-	-	3.207 (7.26)	-	-
Distance to IH-35 * Distance to nearest non-freeway road	-	-	-0.596 (-8.92)	-	-	-	-0.880 (-5.19)	-
Distance to Mopac Freeway	-	-	-	0.089 (4.33)	-	-	-	0.137 (2.41)
Distance to the nearest Road / Distance to the nearest flood plain	-	-	-	0.101 (7.66)	-	-	-	0.240 (5.38)
Parcel lies within Pflugerville City	0.185 (1.87)	0.899 (9.02)	-	-	-0.338 (-1.26)	2.000 (4.52)	-	-
Standard deviation					5.231 (1.434)	-	-	-
Within one kilometer of a school	0.268 (4.56)	-	-	-	0.657 (7.47)	-	-	-
(t=1995) time dummy	-0.145 (-1.21)	-	-	0.349 (3.06)	-0.11 (-1.16)	-	-	0.624 (4.99)
Temporal autocorrelation - ρ	Implicitly restricted to zero				0.367 (3.98)			
Spatial lag - δ	Implicitly restricted to zero				0.449 (5.39)			

The results in Table 4.5 also indicate that parcels farther away from Mopac are more likely to be in an undeveloped state. Mopac is a major expressway connecting the analysis area to the Austin Central Business District (CBD), so it is not surprising that land-owners of parcels located closer to Mopac are more likely to develop their parcels, while land-owners of parcels far away from Mopac may not see the value in developing their land (see Carrión-Flores and Irwin, 2004 and Chakir and Parent, 2009, who also discuss how proximity and access to central metropolitan areas and major roadways can impact land-use decisions). The "push-pull" non-linear effect of distance to the nearest road and distance to the nearest flood plain is clear from the positive coefficient on the ratio of these two variables. Parcels situated within Pflugerville city, according to the MNP model, provide high "net returns" (relative to parcels outside Pflugerville) if invested in residential or commercial land-uses (particularly the latter) rather than being undeveloped or invested in industrial land-use.

However, according to the MNPTS model, on average, parcels within Pflugerville are less likely (relative to parcels outside Pflugerville) to be in residential land-use than being undeveloped or in industrial use. However, there is substantial heterogeneity in this effect, as can be observed from the large estimated standard deviation of the random coefficient on this "Parcel lies within Pflugerville City" variable for the residential land-use alternative. The mean and standard deviation effects on the variable indicate that, for 47.4% of the land-owners of the parcels in the City of Pflugerville, the utility of investing in residential land-use is higher than the utility of leaving the land undeveloped or investing in industrial land-use; for the remaining 52.6% of land-owners of parcels in the City of Pflugerville, the reverse situation holds. Such heterogeneity is a natural result of the tension between the urban amenities (access to retail places and public services such as hospitals) on the one hand that may increase the demand for residential development in already dense residential areas, and the urban "disamenities" (such as traffic congestion effects and air quality problems) on the other hand that may decrease demand for residential development in already dense residential neighborhoods (see Anas et al., 1998; Carrión-Flores and Irwin, 2004 and Irwin and Bockstael, 2002), But, consistent

with the MNP model, the MNPTS model also shows a higher propensity of parcels within Pflugerville City to be invested in commercial land-use than invested in industrial land-use or left undeveloped (see Carrión-Flores et al., 2009). Also, as expected, the proximity to schools is likely to be an incentive to develop the parcel for residential land-use (see Li and Liu, 2007). Finally, the dummy variable for 1995 shows the lower share of parcels in residential land-use and the higher share of parcels in undeveloped land-use in 1995 relative to the other years, as highlighted earlier in Section 4.3.1.

4.3.3.2 Temporal and Spatial Dependency Effects

Temporal dependency (across years) is introduced in the model in the utilities of each alternative for the same land-owner through time-invariant utility preferences and sensitivities to variables (as captured by the random coefficients specification on the constants and the "parcel lies within Pflugerville City" dummy variable in Table 4.5), as well as through the time-varying autoregressive error correlation structure to represent land-owner characteristics that may fade over time (as captured by the autoregressive coefficient ρ). As already indicated in the earlier section, the results show the presence of time-invariant dependency in the utilities for the same land-owner. In addition, Table 4.5 shows a statistically significant and moderate-level autoregressive coefficient of 0.367, indicating the presence of land-owner specific unobserved factors (such as risk averseness or risk acceptance for specific land-use types) that change over time (due to recent events or experiences, or due to lifecycle-related changes). Ignoring these time-varying effects will, in general, lead to inconsistent estimates (due to ignoring the heteroscedasticity generated by these time-varying effects) as well as inefficient estimates (due to ignoring the dependence across the land-use choice occasions of individuals).

The spatial autoregressive parameter in the spatial lag formulation, δ, also turns out to be highly statistically significant with a value of 0.449. This is evidence of the presence of spatial spillover effects caused by dyadic interactions between land-owners of proximately located spatial units. These peer influences are due to strategic or collaborative partnerships between land owners associated with observed and unobserved variables to the analyst, supporting and reinforcing the hypothesis of a spatial lag

formulation to capture spatial dependency in land-use modeling. However, note that this spatial dependence is confined to each of the eastern and western sections of the analysis region (as defined by IH-35), and does not extend to parcels across the two sections. In other words, IH-35 appears to act not simply as a physical barrier, but also as a barrier to peer interactions and influences.

4.3.3.3 Model Selection and Statistical Fit

The MNPTS model is clearly superior to the MNP model, as observed from the statistically significant random coefficients, autoregressive temporal dependence parameter, and the spatial lag parameter. Another way to demonstrate the data fit superiority of the MNPTS model over the MNP model is through the adjusted composite likelihood ratio test (ADCLRT) test. The composite log- likelihood value for the MNP model is -53249.32 (12 parameters estimated) and for the MNPTS model is -51669.8 (17 parameters estimated). The two models may be tested using the adjusted composite likelihood ratio test (AD*CLRT*) statistic (see Pace *et al*., 2011 and Bhat, 2011a). This statistic has a chi-square asymptotic distribution with 5 degree of freedom. The statistic is about 4737, which is higher than the corresponding critical chi-squared value with five degree of freedom at any reasonable level of significance. This demonstrates very strong evidence of temporal dependence and spatial dynamics at play in land-use decisions.

4.3.3.4 Aggregate Elasticity Effects

The estimated parameter coefficients in Table 4.5 provide a sense of the direction of variable effects on the utilities of different land use types. However, these estimated parameters do not directly provide the magnitude of the impact of variables on the probabilities of each land-use category (this is an issue seldom considered in the spatial literature, with many studies simply presenting the parameter results and stopping there). To characterize the magnitude and direction of variable effects on the probabilities, we compute the aggregate-level "elasticity effects" of variables. Specifically, we examine the effects of variables on the expected share of each land-use alternative for the year 2006, given the exogenous variable characteristics of all the 395 parcels. We achieve this by computing the marginal probability of each parcel being in each land-use and aggregating

these probabilities across parcels for each land-use category. The computation of the marginal probability of each parcel being in each land-use is relatively straightforward for the MNP model, so we will focus on the procedure for computing the marginal probabilities from the MNPTS model.

For the MNPTS model, we write the utility function of land-use *i* for the land-owner of parcel *q* as follows (note that the index '*t*' does not appear, since we are focusing on a specific year (2006)):

$$U_{qi} = \delta \sum_{q'} w_{qq'} U_{q'i} + \tilde{\alpha}_{qi} + \boldsymbol{\beta}'_q \mathbf{x}_{qi} + \tilde{\eta}_{qi}\ ;\ \tilde{\alpha}_{qi} = \tilde{a}_i + \breve{\alpha}_{qi},\ \boldsymbol{\beta}_q = \mathbf{b} + \tilde{\boldsymbol{\beta}}_q, \tag{14}$$

where the notation is similar to Section 4.2.1. Next define the following (for ease in presentation, we maintain the same notations as in Section 4.2.1 for the re-defined vectors and matrices):

$$\mathbf{U}_q = (\mathbf{U}'_{q1}, \mathbf{U}'_{q2}, ..., \mathbf{U}'_{qI})' \text{ and } \tilde{\boldsymbol{\eta}}_q = (\tilde{\boldsymbol{\eta}}'_{q1}, \tilde{\boldsymbol{\eta}}'_{q2}, ..., \tilde{\boldsymbol{\eta}}'_{qI})'\ (I \times 1 \text{ vectors}),$$

$$\mathbf{U} = (\mathbf{U}'_1, \mathbf{U}'_2, ..., \mathbf{U}'_Q)' \text{ and } \tilde{\boldsymbol{\eta}} = (\tilde{\boldsymbol{\eta}}'_1, \tilde{\boldsymbol{\eta}}'_2, ..., \tilde{\boldsymbol{\eta}}'_Q)'\ (QI \times 1 \text{ vectors}),$$

$$\breve{\boldsymbol{\alpha}}_q = (\breve{\alpha}_{q1}, \breve{\alpha}_{q2}, ..., \breve{\alpha}_{qI})'\ (\mathrm{I} \times 1 \text{ vector}),\ \breve{\boldsymbol{\alpha}} = \left[(\breve{\alpha}_1)', (\breve{\alpha}_2)', ..., (\breve{\alpha}_Q)'\right]'\ (QI \times 1 \text{ vector}),$$

$$\mathbf{x}_q = (\mathbf{x}_{q1}, \mathbf{x}_{q2}, ..., \mathbf{x}_{qI})'\ (I \times K \text{ matrix}),\ \mathbf{x} = (\mathbf{x}'_1, \mathbf{x}'_2, ..., \mathbf{x}'_Q)'\ (QI \times K \text{ matrix}), \text{ and} \tag{15}$$

$$\tilde{\mathbf{x}} = \begin{bmatrix} x_1 & 0 & 0 & 0. \cdots & 0 \\ 0 & x_2 & 0 & 0 \cdots & 0 \\ 0 & 0 & x_3 & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \ \ldots \ \vdots & \\ 0 & 0 & 0 & 0 \ \ldots \ x_Q & \end{bmatrix}\ (QI \times QK \text{ matrix}), \text{ and}$$

$$\mathbf{S} = \left[\boldsymbol{IDEN}_{\boldsymbol{QI}} - (\delta \boldsymbol{W} \otimes \boldsymbol{IDEN}_{\boldsymbol{I}})\right]^{-1}\ (QI \times QI \text{ matrix}), \tag{16}$$

Then, using other notations as in Section 4.2.1, we may write the following counterpart of Equation (3) for the year 2006:

$$\mathbf{U} = \mathbf{S}\left[\left(\mathbf{1}_Q \otimes \tilde{\mathbf{A}}\right) + \mathbf{x}\mathbf{b} + \breve{\boldsymbol{\alpha}} + \tilde{\mathbf{x}}\breve{\boldsymbol{\beta}} + \tilde{\boldsymbol{\eta}}\right], \tag{17}$$

We simulate the above QI×1-vector $\mathbf{U}$ thousand times using the estimated values of $\delta, \tilde{\mathbf{A}}, \mathbf{b}$, and by randomly drawing 1000 times from the appropriate normal distributions for $\breve{\boldsymbol{\alpha}}, \breve{\boldsymbol{\beta}}$, and $\tilde{\boldsymbol{\eta}}$. Next, we compare the utilities across alternatives for each parcel *q* for

each of the 1000 draws, assign the chosen alternative for each draw, and take the predicted share of each alternative across the 1000 draws to estimate the probability of each parcel being in each land-use alternative. The aggregate share (across parcels) of each land-use type is obtained by aggregating the parcel-level probabilities of each land-use category.

The elasticity computed is a measure of the aggregate percentage change in the aggregate share of each land-use alternative due to a change in an exogenous variable. We also compute the standard errors of the elasticity effects by using 200 bootstrap draws from the sampling distributions of the estimated parameters.[15] For dummy variables, the value of the variable is changed to one for the subsample of intersections for which the variable takes a value of zero, and to zero for the subsample of parcels for which the variable takes a value of one. We then add the shifts in expected aggregate shares in the two subsamples after reversing the sign of the shifts in the second subsample, and compute the effective percentage change in the expected shares across all parcels in the sample due to a change in the dummy variable from 0 to 1. For continuous variables, we increase the value of the variable by 25% for each parcel and compute the percentage change in the expected shares.

The elasticity effects and their standard errors are computed for the MNP model and the MNPTS model, and are presented in Table 4.6. The effects (and their standard errors in parenthesis) are presented for the six scenarios listed in the table. The first entry in the table indicates that, on average, a parcel that is within 350 meters from IH-35 is about 35.1% less likely to be in residential land-use relative to a parcel that is beyond 350 meters of IH-35. Similarly, the entry in the first column and second row suggests that a parcel that is 25% farther away from IH-35 than another parcel is about 1.2% more likely to be in residential land-use than the closer-to-IH35 parcel.

[15] For ease in computation, we however fix the spatial lag parameter δ in the bootstrapping, so that we do not have to compute the matrix **S** for each bootstrap draw (the matrix **S** entails a high-dimensional matrix inversion).

Table 4.6. Aggregate-Level Elasticity Effects of the MNP and MNPTS Models (standard error in parenthesis)

Scenario	Residential			Commercial			Industrial			Undeveloped		
	MNP	MNPTS	p†	MNP	MNPTS	p	MNP	MNPTS	p	MNP	MNPTS	p
A change from the parcel being farther than 350 meters from IH-35 to within 350 meters from IH-35	-35.1 (3.0)	-67.3 (6.5)	0.00	382.5 (45.9)	806.2 (129.1)	0.00	-37.8 (3.2)	-75.2 (7.6)	0.00	-32.7 (2.7)	-54.6 (5.7)	0.00
A 25% increase in the dist. to IH-35, but only for those parcels farther than 350 meters from IH-35,	1.2 (0.1)	3.4 (0.4)	0.00	0.9 (0.2)	1.8 (0.5)	0.08	-11.2 (0.6)	-22.1 (2.2)	0.00	1.1 (0.1)	1.0 (0.2)	-*
A 25% increase in the distance to the nearest flood plain	1.9 (0.3)	3.0 (0.5)	0.09	1.4 (0.3)	2.2 (0.5)	0.13	1.2 (0.2)	1.8 (0.5)	-	-2.5 (0.4)	-3.8 (0.6)	0.06
A 25% increase in distance to the nearest road and a 25% decrease in the distance to the nearest flood plain	-6.6 (0.7)	-9.8 (1.4)	0.03	-5.0 (0.7)	-6.7 (1.3)	-	-4.0 (0.5)	-6.3 (1.4)	0.15	8.5 (0.9)	12.3 (1.7)	0.04
A switch of the parcel location from Pflugerville to outside Pflugerville	-5.1 (9.5)	-15.3 (11.1)	-	253.6 (41.7)	380.4 (143.0)	-	-41.8 (7.7)	-24.3 (19.4)	-	-29.1 (6.2)	-34.9 (7.6)	-
A switch of the parcel location from being farther than one km. from the closest school to being closer than one km. from the closest school	34.7 (9.4)	89.7 (17.3)	0.01	-17.5 (3.5)	-22.2 (3.5)	-	-19.3 (4.1)	-64.5 (5.4)	0.00	-14.8 (3.1)	-21.1 (2.9)	0.14

† p value of the difference

*A '-' implies that the difference is not statistically significant even at the 0.2 level of significance

Other entries may be similarly interpreted. The last sub-column within each alternative column provides the p-value for the difference in elasticity estimates from the MNP and MNPTS models. A '-' in this column implies that the difference is not statistically significant even at the 0.2 level of significance.

The elasticity effects of both the MNP and MNPTS models are in the same direction for all variables, and are consistent with the discussions in the previous section. However, it is clear than the elasticity effects from the MNPTS model are generally higher in magnitude than those from the MNP model, a consequence of the "spillover" effects in the MNPTS model that causes a spatial multiplier effect. Specifically, a change in a variable for one parcel influences the utilities of the land-use alternatives of other parcels, which then have a "circular" influence back on the utilities of the land-use alternatives for the parcel for which a variable has been changed. This "circular" influence is reinforcing because of the positive spatial lag parameter, which implies the spatial multiplier effect (this spatial multiplier effect is captured by the $\mathbf{S}$ matrix in Equation (17)). The MNP model ignores the presence of such spatial multiplier effects, and assumes that a change in a variable at one parcel impacts only the land-use at that parcel.

The difference in the elasticity effects between the MNP and MNPTS models are, for the most part, statistically significant. Thus, the higher MNPTS-predicted positive effects of a parcel being within 350 meters of IH-35 (rather than being beyond 350 meters of IH-35) on the probabilities of the parcel being in commercial land-use, and the higher MNPTS-predicted negative effect of a parcel being within 350 meters of IH-35 (rather than being beyond 350 meters of IH-35) on the probability of the parcel being in non-commercial land-uses, are all highly statistically significant. Similarly, the differential effects (between the MNP and MNPTS models) of the continuous distance from IH-35 (second variable in Table 4.6) on the probabilities of the residential and industrial land-uses are highly statistically significant, while the differential effects on the probabilities of commercial land-use are also quite statistically significant. Other differences and their p-values may be similarly extracted from Table 4.6. The one

variable for which there is no statistically significant difference in the MNP and MNPTS elasticity effects is for the Pflugerville City variable (see the last but one row of the table). For this variable, while the elasticity effects are indeed higher from the MNPTS model, the heterogeneity in the utility for the residential land-use type leads to a tempering of the effects on the utilities of other alternatives, which counteracts the spatial multiplier effect. The heterogeneity also leads to higher standard errors for the elasticity estimates. In combination, the tempered effects on elasticities and the higher standard errors lead to less statistically significant differences. But, overall, there are statistically significant differences in elasticity predictions between the MNP and MNPTS models, highlighting the predictive differences between the two models and, in general, the under-estimations of the magnitudes of variable effects from the MNP model.

The elasticity effects from the continuous variables (such as the continuous distance to IH-35) are not directly comparable to those from the dummy variables (such as whether or not the parcel is within 350 meters of IH-35). However, the results identify closeness to IH-35 (whether within a 350 meters band of IH-35 or not), Pflugerville location, and proximity to schools as the dominant variables impacting the land-use type of a given parcel.

4.4 CONCLUSION

This study has proposed a new econometric approach to specify and estimate a model of land-use change, based on the now rich theoretical literature on land use conversion decisions made by economic agents to maximize net returns. At a methodological level, the study has formulated and estimated a multi-period multinomial probit model, accounting for time-varying and time-stationary inter-temporal dependencies as well as a spatial lag structure across observation units. The model also accommodates spatial heterogeneity. The inference methodology used is the maximum approximate composite marginal likelihood (MACML) approach. The study has modeled the land-use type of parcel-level spatial units in an area north of the City of Austin in Texas. In doing so, the emphasis has been on better linking the quantitative (but aspatial or highly stylized spatial effects) perspective for land-use analysis that dominates the economic literature

with the qualitative (but richer spatial dynamics and heterogeneity) perspective for land-use analysis that is quite prevalent in the ecological literature. The empirical results indicate the presence of statistically significant time-invariant and time-varying land-owner-specific unobserved factors as well as the presence of spatial spillover effects caused by dyadic interactions between land-owners of proximately located spatial units. Ignoring these dependencies and dynamics will, in general, lead to inconsistent and inefficient estimates of parameter effects. This is highlighted by computing the elasticity effects of variables, which indicates that the model that accommodates temporal dependencies and spatial dynamics predicts magnitude effects that are statistically significantly different from the model that ignores these effects. Important determinants of land-use type include proximity to highways and other roadways, distance from flood plains, parcel location in the context of existing development, and distance from schools. The results also suggest that major transportation roadways can act not only as physical separators of land areas, but also as a barrier to peer interactions and influences. To conclude, the model structure and inference approach proposed in this study should be applicable in a wide variety of fields where social and spatial interactions (or dyadic interactions) between decision-makers lead to spatial multiplier and spillover effects in the choices of the decision-makers.

CHAPTER 5: A NEW APPROACH TO SPECIFY AND ESTIMATE NON-NORMALLY MIXED MULTINOMIAL PROBIT MODELS

The material in this chapter is drawn substantially from the following published paper.

Bhat, C.R., and R. Sidharthan (2012), A New Approach to Specify and Estimate Non-Normally Mixed Multinomial Probit Models. *Transportation Research Part B* 46(7), 817-833[16]

5.1 INTRODUCTION

Econometric discrete choice analysis is an essential component of studying individual choice behavior and is used in many diverse fields to model consumer demand for commodities and services. The decision principle used in almost all discrete choice models corresponds to utility maximization, which is based on the Lancastrian (1971) notion of the assignment of a composite utility to each alternative in the choice set (based on alternative and individual attributes) followed by the choice of the alternative with the highest utility. Further, since the analyst does not observe all individual and context-related factors that contribute to choice decisions, one or more stochastic elements (or random error terms) are introduced in the utility of alternatives. Different ways of introducing the stochastic elements lead to different discrete choice model structures. Thus, consider a cross-sectional choice situation with a single choice occasion per individual, and assume independence among the choice behaviors of individuals.[17] Then, the simplest model form, corresponding to the multinomial logit (MNL) model introduced by Luce and Suppes (1965) and McFadden (1974), assumes a *single composite* independently and identically distributed or IID (across alternatives) random utility error term with a Gumbel (or Type I extreme-value) distribution. This leads to the

[16] The author of this dissertation collaborated with the coauthor on the methodological and technical aspects of the paper.

[17] The use of a cross-sectional choice situation with independence across individual decision-maker choices is simply for exposition convenience in this introduction section.

simple and elegant MNL model form, but also leaves the model form saddled with the familiar independence from irrelevant alternatives (IIA) property. Maintaining a single composite Gumbel error term in utilities, while relaxing the independence assumption (across alternatives), moves the model form from the multinomial logit to the generalized extreme-value (GEV) class of models proposed by McFadden (1978). On the other hand, relaxing the identically distributed assumption (across alternatives) with the Gumbel distribution assumption leads to the Heteroscedastic Extreme Value (HEV) model form proposed by Bhat (1995). Finally, still maintaining a single composite error term but now with a normal distribution, when combined with relaxation of the independence and/or identical distribution assumptions, generates the multinomial probit (MNP) model form originally proposed by Hausman and Wise (1978) and Daganzo (1979). Of these model forms, the MNP form allows the most flexible error covariance structures (up to certain limits of identifiability; see Train, 2009, Chapter 5), though it also entails more estimation effort since it requires the evaluation of a multidimensional normal orthant probability function with an $(I-1)$ dimensional integral in the general case (where I is the number of alternatives).

A substantial amount of the early theoretical developments in discrete choice modeling was focused on a single composite error term. Over the past decade and a half, attention has shifted more toward the use of multiple error terms through the introduction of a mixing random distribution structure in the utility function of alternatives that is independent of the kernel error term. Essentially, the mixing structure superimposes additional stochastic terms over the "kernel" error term discussed in the previous paragraph. There are several reasons for this shift toward mixing structures. First, in a cross-sectional context, it is very plausible that there are unobserved variations across individuals in the sensitivity to relevant exogenous attributes (such as differential sensitivity due to unobserved factors to travel time and travel cost in a travel mode choice model). Ignoring these variable-specific stochasticity effects and instead using a single composite error term in the utility function will, in general, lead to inconsistent coefficient estimates and trade-off estimates, as well as incorrect substitution patterns

across alternatives (see Bhat, 1997a).[18] A second reason for the increasing use of mixing structures is that they provide the ability to introduce heteroscedasticity across utilities in the closed-form GEV models through an error-components specification, as discussed in Train (2009). It also provides the ability to generate correlation across alternatives through an error-components specification. The use of a mixing structure over the closed-form GEV kernel-based model can then essentially achieve any desired covariance pattern. At the same time, and especially when the number of alternatives far exceeds the number of mixing random terms needed to capture the "true" covariance pattern, the maximum simulated likelihood (MSL) estimation of the mixed GEV model is generally much easier and faster than a non-mixed MNP model (see Bhat *et al*., 2008 and Train, 2009 for detailed discussions). A third reason for using mixing structures is that, when using GEV-based kernels, mixing structures enable the introduction of error dependencies across the choice occasions of the same decision-maker in panel or repeated choice contexts (see Li *et al*., 2010). Even when using an MNP kernel, the mixing structure can provide substantial econometric and computational efficiency to capture panel effects. Further, the mixing approach is almost identical when dealing with cross-sectional choice data or panel data, and poses no conceptual and likelihood estimation coding differences.

There is yet another reason to consider a mixing approach in discrete choice modeling. This has to do with explicitly specifying the random mixing distribution on

[18] There are a few exceptions to this rule, one of which is when an MNP kernel error term is mixed with normally distributed random coefficients. Assuming the usual linear-in-parameters utility functional form, the net effect is that the combination of variable-specific random terms and the kernel error term can be recast back into an MNP utility form with a single composite error term (due to the closure property of the normal distribution under affine transformations -- a linear transformation followed by a translation). That is, the marginal distribution of utility obtained by integrating out the normal mixing distribution puts the utility back into a normal distribution form. In fact, this was the genesis of Hausman and Wise's MNP model formulation, in which the "composite" error terms of the alternatives have a covariance matrix that is parameterized based on the mixing structure. However, this kind of affine closure is not achieved with GEV or HEV kernel models. Further, closure is also not generally achieved with a non-normal mixing distribution with the MNP "kernel", except in a special case which is exploited in this study.

variables in a way that is consistent with theoretical notions. In fact, the ability to do so is critical to the observation made by McFadden and Train (2000) that the mixed multinomial logit model is capable of approximating any random utility maximization model. Thus, for example, one may want to consider bounded distributions (such as a log-normal distribution or a Rayleigh distribution) for cost and time coefficients in a travel mode choice model, so that the coefficients on these variables are bounded at the upper end. On the other hand, the coefficients on some other variables may be appropriately considered as being unbounded. Further, there are several types of continuous distributions that may be used to capture the profile of population sensitivity to variables.[19] In the context of continuous mixing distributions, the normal distribution has been used quite extensively in the past. However, several studies (see, for example, Amador *et al.*, 2005, Train and Sonnier, 2005, Hensher *et al.*, 2005, Fosgerau, 2005, Greene *et al.*, 2006, Balcombe *et al.*, 2009, and Torres *et al.*, 2011) have underscored the potentially serious mis-specification consequences (in terms of theoretical considerations, data fit, as well as trade-off evaluations) of using the normal distribution. In particular, the symmetric nature of the normal distribution, when combined with mean values that may not be too far away from zero, implies that a significant fraction of individuals may

[19] Note here that discrete distributions may also be used for the mixing. If the mixing vector is assumed to take M possible value states with state-specific probabilities, this leads to the familiar latent class model used in marketing (see Kamakura and Russell, 1989, Chintagunta *et al.*, 1991) and transportation (see Bhat, 1997b, Greene and Hensher, 2003, Hess *et al.*, 2007, and Train, 2008). On the other hand, if a discrete distribution is considered separately for each individual random coefficient, this is essentially a non-parametric distribution (see Bastin *et al.*, 2010, Cherchi *et al.*, 2009, Fosgerau, 2006). However, the use of a continuous distribution dominates the literature, at least in part because it offers efficiency in the number of mixing distribution parameters to be estimated. Further several studies that have compared discrete distribution methods with continuous distributions have not found a clear pattern of which of the two approaches is superior (see, for instance, Greene and Hensher, 2003, Birol *et al.*, 2006, and Hynes *et al.*, 2008). Some recent studies have also considered a combination of discrete and continuous distributions for the mixture in the form of a mixture of normal distributions (see Campbell *et al.*, 2010), though such mixtures of normal distributions have some of the same problems as the simple normal distribution (as discussed subsequently).

have an unexpected sign on variables (such as a positive coefficient on cost or time). For instance, Train and Sonnier (2005), in their analysis of vehicle choice, found that 22% of the population preferred vehicles with a higher purchase price, and 37% of the population preferred vehicles with a higher operating cost, when they used a normal distribution for the cost coefficients. On the other hand, when Train and Sonnier used a log-normal distribution and a bounded Johnson's SB distribution for the cost coefficients, such results were avoided and they also obtained better data fits. Finally, another issue with using normally distributed cost and other coefficients is that this leads to a breakdown of the WTP calculation because the moments of the ratio of two normally distributed random terms do not exist (see Cedilnik *et al*., 2006, Daly *et al*., 2011).

As indicated already, there have been several earlier studies that have successfully estimated non-normal distributions for the mixing distribution. All of these studies use a multinomial logit model kernel over which mixing is specified. However, the general experience has been that, even when successful, such estimations take a longer time for convergence (relative to normal distributions). This is particularly so for asymmetric distributions with long tails, such as the log-normal distribution. Further, in some cases, the maximum simulated likelihood (MSL) of models with non-normal mixing fails due to numeric/computational problems. It is not uncommon to see researchers consider non-normal distributions only to eventually revert to the use of a normal distribution (see, for example, Bartels *et al*., 2006 and Small *et al*., 2005). In addition to these problems specific to the use of non-normal distributions, MSL inference techniques can have other limitations, including a rapid degradation in accuracy as the number of dimensions of mixing increases, and problems with the accuracy (or lack thereof) of the covariance matrix of the estimator. These issues may be traced back to the use of a simulation approach to evaluate the log-likelihood function, which leads to a highly nonlinear and non-smooth second derivatives surface of the log-simulated likelihood function.

Recently, Bhat (2011a) proposed an alternative maximum approximate composite marginal likelihood (MACML) inference approach to estimate the multinomial probit (MNP) model. His basis for preferring an MNP kernel rather than a

multinomial logit or GEV kernel originates from several considerations. First, in cases such as a spatial analysis where the utility of spatial alternatives are correlated based on proximity, or in situations where the utility of individuals for alternatives have a spatial dependency component based on the usual spatial error/lag formulations used in spatial econometrics (see Anselin, 1988), the resulting parametric covariance structure across alternatives or across decision-makers is simply infeasible or extremely inefficient to incorporate with a mixing approach over a restrictive Gumbel kernel covariance surface. Second, when a normal mixing distribution is used, the resulting "mixed MNP" model collapses back to an MNP model due to the closure property of the normal distribution under affine transformations. This, along with the MACML inference procedure, implies the need only to evaluate univariate and bivariate cumulative normal distribution function evaluations, regardless of the number of alternatives or the number of choice occasions per individual or the nature of social/spatial dependence structures. Further, the MACML procedure uses an *analytic approximation* method rather than a *simulation evaluation* method to evaluate the multivariate normal cumulative distribution function, which improves the ability to accurately and precisely recover the parameters and their covariance matrix estimates (because of the smooth nature of the first and second derivatives of the approximated analytic log-likelihood function). The net result is that the MNP kernel with the MACML inference approach leads to substantial computational gains compared to the MSL estimation of normally-mixed MNL and GEV models, as well as enables estimation in cases where the MSL estimation of mixed MNL and GEV approaches are simply infeasible.

One problem, however, with Bhat's MACML approach as it stands is that it is only applicable to the normally-mixed case. However, as discussed earlier, a normal mixing distribution may not be appropriate in several cases. What is needed then is a model that is able to include both a general covariance kernel structure as well as non-normal mixing, while also still being able to be estimated using the MACML approach. This is the objective of this chapter. Specifically, we introduce the use of a multivariate skew-normal distribution function for mixing with an MNP kernel model. The skew-

normal distribution, considered by O'Hagan and Leonard (1976) and formalized by Azzalini (1985) for the univariate case, has been extended to the multivariate case by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). Since these initial contributions, more research on different types of multivariate generations of the skew-normal distribution and their properties have been undertaken (see Gonzalez-Farias *et al.*, 2004, Arellano-Valle and Genton, 2005, Gupta *et al.*, 2004, Arellano-Valle and Azzalini 2006, 2008, Azzalini, 2011). As discussed later, the multivariate skew normal (MSN) distribution retains several attractive properties of the multivariate normal distribution, and an MNP kernel model mixed with this distribution also lends itself nicely to estimation using the MACML approach. At the same time, the MSN distribution is tractable, parsimonious in parameters that regulate the distribution and its skewness, and includes the normal distribution as a special interior point case. It also is a very flexible unimodal density structure that allows a "seamless" and "continuous" variation from normality to non-normality, and can replicate a variety of smooth unimodal density shapes with tails to the left or right as well as with a high modal value (sharp peaking) or low modal value (flat plateau). The asymmetry accruing from the skewness of the distribution also can allow the density to be pretty much confined to the positive (or negative) half-line. In this sense, it includes a likeness of the log-normal density function as a special case, but with tails that are thin as in the normal density function (which makes estimation easier than in the log-normal case). Despite these desirable properties, there has been little explicit consideration of the skew normal distribution for random terms even in the linear regression field with continuous observations (but see Jara *et al.*, 2008, Meintanis and Hlavka, 2010, and Molenaar *et al*., 2010), and there has been no consideration whatsoever of this distribution in the discrete choice field.[20]

[20] However, it should be noted that the skew normal distribution has appeared implicitly in the context of such models as the stochastic frontier model (see Aigner *et al*., 1977) and in other studies involving the study of truncated normal variables (for example, Birnbaum, 1950 and Weinstein, 1964). This is because one of the stochastic representations of a skew-normally distributed variable happens to be as the convolution of a normal variable and a half-normal variable. However, the explicit use of the skew-normal

The rest of this chapter is structured as follows. The next section discusses the fundamental structure and properties of the univariate and multivariate skew normal distributions. Section 5.3 presents the model framework and estimation procedure for the proposed skew-normally mixed MNL model. Section 5.4 undertakes a simulation exercise to assess the ability of the proposed model to recover underlying parameters. Finally, Section 5.5 summarizes the key findings of the chapter.

5.2 THE SKEW-NORMAL DISTRIBUTION

The literature on the skew-normal distribution is quite vast, but also scattered. In this section, we compile and present all the most relevant properties of the distribution in the context of application for mixed MNP models. The section begins with a characterization of the univariate skew-normal distribution and then proceeds to the more relevant case of the multivariate skew-normal distribution.

5.2.1 The Univariate Skew-Normal Distribution

A random variable Y is labeled as being skew-normally distributed with a location parameter ξ $(\xi \in \Re)$, a scale parameter ω $(\omega > 0)$, and a shape parameter α $(\alpha \in \Re)$ if its probability density function is as follows:

$$f(y;\xi,\omega^2,\alpha) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left\{\alpha\left(\frac{y-\xi}{\omega}\right)\right\}, \tag{1}$$

where $\phi(.)$ and $\Phi(.)$ represent the standard normal density and cumulative distribution function, respectively. When $\alpha = 0$, the density collapses to that of a normal distribution with mean and variance parameters of ξ and ω^2, respectively. Setting Y = $\xi + \omega Z$, we obtain a standardized version of the probability density function of the skew-normal distribution (corresponding to the density function of Z that has a location parameter of 0 and scale parameter of 1) given by $\tilde{\phi}(z;\alpha) = 2\phi(z)\Phi(\alpha z)$. The density function for Y in Equation (1) may be written in terms of the standard density function as $\omega^{-1}\tilde{\phi}(z;\alpha)$,

---

as a distributional assumption for one or more random terms, as in the current study, has seen little consideration in the econometric field.

where $z = \omega^{-1}(y - \xi)$. Appendix A.1 presents the moment generating function and the moments of the standardized skew-normal distribution (SSN).

An important stochastic representation for *Z* that is useful for random generation from the SSN distribution is obtained using a conditioning mechanism. Specifically, consider two bivariate normally distributed variables M_1 and M_2:

$$\begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right). \tag{2}$$

Then, $Z = M_2 \,|\, (M_1 > 0)$ has the SSN density function $\tilde{\phi}(z;\alpha)$, where the relationship between ρ and α is as follows: $\rho = \frac{\alpha}{\sqrt{1+\alpha^2}}$ (see Appendix A.2 for a derivation). Using this conditioning mechanism, the cumulative distribution function for Z may be obtained as follows:

$$\begin{aligned} P(Z < z) &= \tilde{\Phi}(z;\, \alpha) = \frac{P(M_1 > 0, M_2 < z)}{P(M_1 > 0)} \\ &= 2\,P(-M_1 < 0, M_2 < z) = 2\Phi_2(0, z, -\rho); \rho = \frac{\alpha}{\sqrt{1+\alpha^2}}. \end{aligned} \tag{3}$$

Thus, the cumulative SSN distribution function may be written in terms of a bivariate cumulative standard normal distribution function, and the cumulative distribution function for the non-standardized skew-normally distributed variable *Y* may be obtained as:

$$P(Y < y) = \tilde{\Phi}\left(\frac{y-\xi}{\omega};\, \alpha\right) = 2\Phi_2\left(0, \frac{y-\xi}{\omega}, -\rho\right); \rho = \frac{\alpha}{\sqrt{1+\alpha^2}}. \tag{4}$$

For the extension to the multivariate skew-distribution, and especially for use with the multinomial probit model, an alternate parameterization of *Z* (referred to by Arellano-Valle and Azzalini, 2006 as the unified skew-normal variable) will be helpful. This is based on the conditioning mechanism discussed above. In this alternate parameterization, the univariate SSN density function is written as $\tilde{\phi}(z;\rho)$ and the univariate cumulative distribution function is written as $\tilde{\Phi}(z;\rho)$, with ρ replacing α.

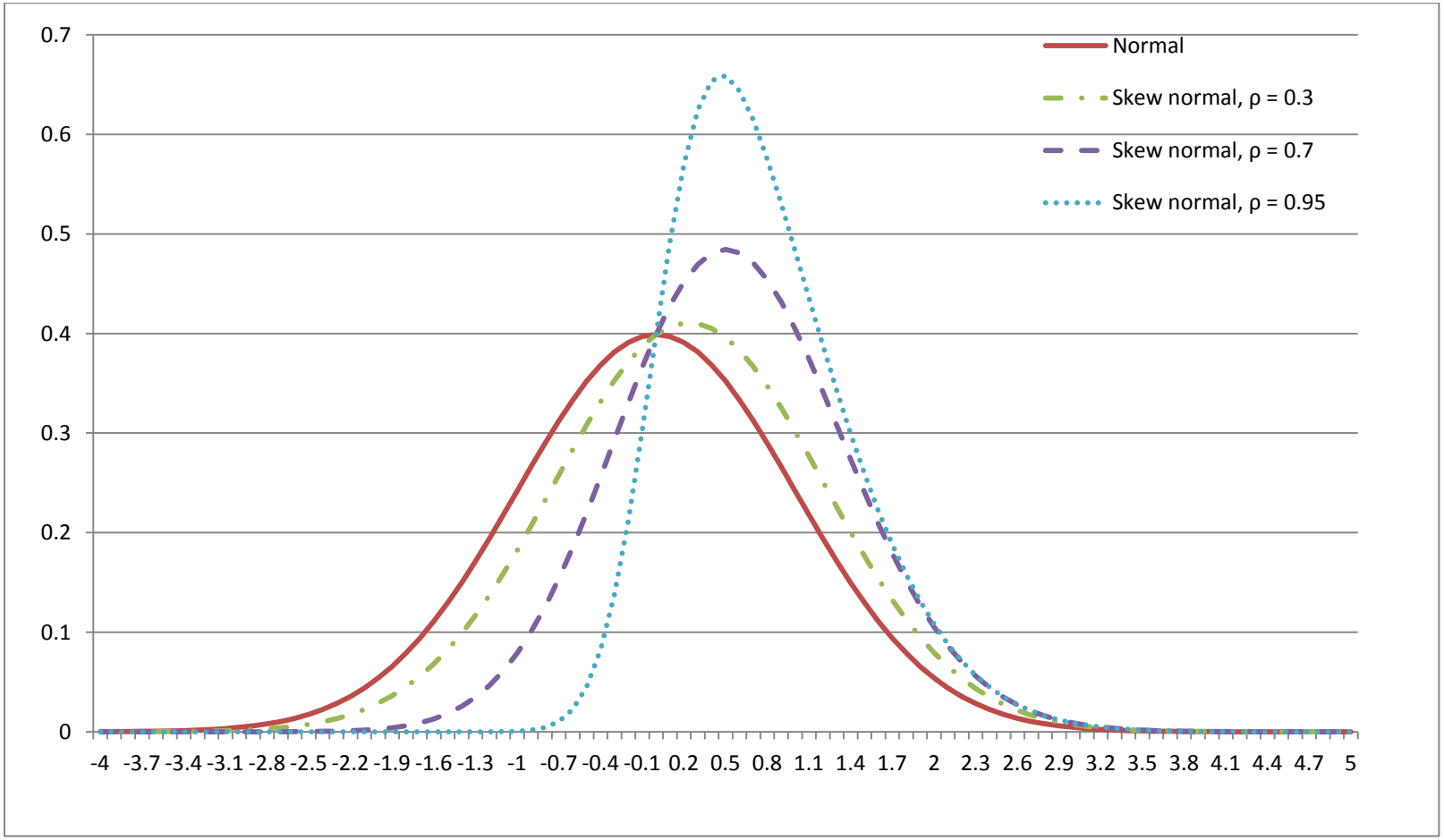

Figure 5.1. Shape of the SSN density function for a number of positive values of ρ

Figure 5.1 shows the shapes of the normal density function (solid line) and the SSN density functions for three positive values of ρ (the plots are mirrored across the y-axis for negative values of ρ). As the value of the shape parameter ρ increases, the skewness of the distribution increases and the density shows sharper peaking. As $\rho \to 1$, the SSN density tends toward a half-normal density function. Note also that, as the shape parameter increases, the right skewness increases not because the extreme right tail gets longer but because the left tail becomes shorter and shorter (relative to the normal distribution). This is a desirable property in the likelihood convergence of mixed models, and is unlike the log-normal distribution whose right tail gets very long rapidly as the variance of the distribution increases.

5.2.2 The Multivariate Skew-Normal Distribution Function

There are several multivariate versions of the skew-normal distribution in the literature (see Arellano-Valle and Azzalini, 2006 for a discussion of these many variants, and a unified treatment of these). All of these share several properties similar to the multivariate normal distribution. In this study, we select the multivariate skew distribution version originally proposed by Azzalini and Dalla Valle (1996) for a number of reasons. This version is efficient in the number of additional parameters to be estimated, allows independence between skew-normally distributed and normally-distributed elements in a multivariate vector (useful in selectively imposing skew-normality only on certain coefficients), is closed under any affine transformation of the skew-normally distributed vector (is the key to the MACML estimation of the MNP model), and is closed under the sum of independent skew-normally distributed and normally distributed vectors of the same dimensions (is the key to non-normally mixing distributions superimposed on an MNP kernel). As importantly, the cumulative distribution function of a *D*-variate skew normally distributed variable of the Azzalini and Dalla Valle type requires only the evaluation of a $(D+1)$-dimensional multivariate cumulative normal distribution function.

Consider a multivariate skew-normally (MVSN) distributed random variable vector $\mathbf{Y} = (Y_1, Y_2, Y_3, \ldots, Y_D)'$ with a $(D \times 1)$-location parameter vector ξ ($\xi \in \Re^D$), and

a $(D\times D)$-symmetric positive-definite covariance matrix $\boldsymbol{\Omega}$. Let the correlation matrix corresponding to $\boldsymbol{\Omega}$ be $\boldsymbol{\Omega}^*$, and let $\boldsymbol{\omega}$ be a $(D\times D)$-diagonal matrix formed by the standard deviations of $\boldsymbol{\Omega}$ (ω_j is the *j*th diagonal element of the matrix $\boldsymbol{\omega}$). Then, we may write: $\boldsymbol{\Omega}^* = \boldsymbol{\omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\omega}^{-1}$. Setting $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, we obtain a standardized version of the multivariate probability density function of the skew-normal distribution (corresponding to the density function of $\mathbf{Z}$ that has a location parameter of $\mathbf{0}$ and a correlation matrix $\boldsymbol{\Omega}^*$). As in the univariate case, it can be shown that the random variable $\mathbf{Z}$ is obtained through a latent conditioning mechanism on a $(D+1)$-variate normally distributed vector $\tilde{\mathbf{M}} = (\tilde{M}_1, \tilde{\mathbf{M}}_2')'$, where $\tilde{M}_1$ is a latent (1×1)-vector and $\tilde{\mathbf{M}}_2$ is a $(D\times 1)$-vector:

$$\begin{pmatrix} \tilde{M}_1 \\ \tilde{\mathbf{M}}_2 \end{pmatrix} \sim N_{d+1}\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \boldsymbol{\Omega}_+^*\right), \text{ where } \boldsymbol{\Omega}_+^* = \begin{pmatrix} 1 & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \boldsymbol{\Omega}^* \end{pmatrix}. \tag{5}$$

$\boldsymbol{\rho}$ is a $(D\times 1)$-vector, each of whose elements must lie between –1 and +1. The matrix $\boldsymbol{\Omega}_+^*$ is also a positive-definite correlation matrix. Then, $\mathbf{Z} = \tilde{\mathbf{M}}_2 \mid (\tilde{M}_1 > 0)$ has the standard multivariate skew-normal (SMVSN) density function shown below:

$$\tilde{\phi}_D(\mathbf{z}; \boldsymbol{\Omega}_+^*) = 2\phi_D(\mathbf{z}; \boldsymbol{\Omega}^*)\,\Phi(\boldsymbol{\alpha}'\mathbf{z}), \text{ where } \boldsymbol{\alpha} = \frac{(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{\rho}}{\left(1 - \boldsymbol{\rho}'(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{\rho}\right)^{1/2}}. \tag{6}$$

where $\phi_D(.)$ and $\Phi(.)$ represent the standard multivariate normal density function of *D* dimensions and the standard univariate cumulative distribution function, respectively. We write $\mathbf{Z} \sim \text{SMVSN}(\boldsymbol{\Omega}_+^*)$. The probability density function of the random variable $\mathbf{Y}$ [$\mathbf{Y} \sim \text{MVSN}(\boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\Omega}_+^*)$] may be written in terms of the SMVSN density function above as:

$$f_D(\mathbf{y}; \boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\Omega}_+^*) = \left(\prod_{j=1}^{D} \omega_j\right)^{-1} \tilde{\phi}_D(\mathbf{z}; \boldsymbol{\Omega}_+^*), \text{ where } \mathbf{z} = \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}). \tag{7}$$

The moment generating function of $\mathbf{Z}$ and its first three moments are presented in Appendix A.3.

The cumulative distribution function for $\mathbf{Z}$ may be obtained as:

$$P(\mathbf{Z} < \mathbf{z}) = \widetilde{\Phi}_D(\mathbf{z}; \boldsymbol{\Omega}^*_+) = 2\Phi_{D+1}(\mathbf{0}, \mathbf{z}, \boldsymbol{\Omega}^*_-);\ \boldsymbol{\Omega}^*_- = \begin{pmatrix} 1 & -\boldsymbol{\rho}' \\ -\boldsymbol{\rho} & \boldsymbol{\Omega}^* \end{pmatrix}. \tag{8}$$

The corresponding cumulative distribution function for $\mathbf{Y}$ is:

$$P(\mathbf{Y} < \mathbf{y}) = \widetilde{\Phi}_D\left(\boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}); \boldsymbol{\Omega}^*_+\right) = 2\Phi_{D+1}\left(0, \boldsymbol{\omega}^{-1}(\mathbf{y} - \boldsymbol{\xi}), \boldsymbol{\Omega}^*_-\right). \tag{9}$$

The close correspondence with the normal distribution leads to several desirable properties of the multivariate skew-normal (MVSN) distribution. The ones that are key to the proposal in this study to use the MSN distribution for mixing in MNP models are listed and discussed below.

Property 1:

The sum of a MVSN distributed vector $\mathbf{Y}$ (of dimension $D \times 1$) [$\mathbf{Y} \sim \text{MVSN}(\boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\Omega}^*_+)$] and an independently distributed multivariate normally (MVN) distributed vector $\mathbf{W}$ (also of dimension $D \times 1$) [$\mathbf{W} \sim \mathbf{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$] is still MVSN distributed:

$$\mathbf{Y} + \mathbf{W} \sim \text{MVSN}(\boldsymbol{\xi} + \boldsymbol{\mu}, \widetilde{\boldsymbol{\omega}}, \widetilde{\boldsymbol{\Omega}}^*_+), \quad \text{where} \quad \widetilde{\boldsymbol{\Omega}}^*_+ = \begin{pmatrix} 1 & \widetilde{\boldsymbol{\rho}}' \\ \widetilde{\boldsymbol{\rho}} & \widetilde{\boldsymbol{\Omega}}^* \end{pmatrix}, \widetilde{\boldsymbol{\Omega}}^* = (\widetilde{\boldsymbol{\omega}})^{-1}\widetilde{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{\omega}})^{-1}, \widetilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega} + \boldsymbol{\Sigma},$$

$\widetilde{\boldsymbol{\rho}} = (\widetilde{\boldsymbol{\omega}})^{-1}\boldsymbol{\omega}\boldsymbol{\rho}$, and $\widetilde{\boldsymbol{\omega}}$ is the diagonal matrix of standard deviations of $\widetilde{\boldsymbol{\Omega}}$.

Proof: There are several ways to prove this property, but perhaps the easiest is to use the moment generating functions of $\mathbf{Y}$ and $\mathbf{W}$. Specifically, we have (from Appendix A.3):

$$\begin{aligned} M_{Y+W}(\mathbf{t}) = M_Y(\mathbf{t}) \times M_W(\mathbf{t}) &= \left[2\exp\left(\boldsymbol{\xi}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}\right)\Phi(\boldsymbol{\rho}'\boldsymbol{\omega}\,\mathbf{t})\right] \times \left[\exp\left(\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\right)\right] \\ &= \left[2\exp\left((\boldsymbol{\xi}' + \boldsymbol{\mu}')\mathbf{t} + \frac{1}{2}\mathbf{t}'(\boldsymbol{\Omega} + \boldsymbol{\Sigma})\mathbf{t}\right)\Phi(\widetilde{\boldsymbol{\rho}}'\widetilde{\boldsymbol{\omega}}\,\mathbf{t})\right], \text{ where } \widetilde{\boldsymbol{\rho}} = (\widetilde{\boldsymbol{\omega}})^{-1}\boldsymbol{\omega}\boldsymbol{\rho}. \end{aligned} \tag{10}$$

The above expression is once again in the MVSN moment generating form in Appendix (A.3), from which the property is proved.

Property 2:

The affine transformation of the MVSN distributed vector $\mathbf{Y}$ (of dimension $d \times 1$) $[\mathbf{Y} \sim \text{MVSN}(\xi, \omega, \Omega_+^*)]$ as $\mathbf{a} + \mathbf{BY}$, where $\mathbf{B}$ is a $(h \times d)$ matrix is also a MVSN distributed vector of dimension $h \times 1$:

$\mathbf{a} + \mathbf{BY} \sim \text{MVSN}(\mathbf{a} + \mathbf{B}\xi, \tilde{\omega}, \tilde{\Omega}_+^*)]$,

where $\tilde{\Omega}_+^* = \begin{pmatrix} \mathbf{1} & \tilde{\rho}' \\ \tilde{\rho} & \tilde{\Omega}^* \end{pmatrix}$, $\tilde{\Omega}^* = (\tilde{\omega})^{-1}\tilde{\Omega}(\tilde{\omega})^{-1}$, $\tilde{\Omega} = \mathbf{B}\Omega\mathbf{B}'$,

$\tilde{\rho} = (\tilde{\omega})^{-1}\mathbf{B}\omega\,\rho$, and $\tilde{\omega}$ is the diagonal matrix of standard deviations of $\tilde{\Omega}$.

Proof: The moment generating function of $\mathbf{a} + \mathbf{BY}$ may be written as:

$$M_{\mathbf{a}+\mathbf{BY}}(\mathbf{t}) = M_{\mathbf{Y}}(\mathbf{a} + \mathbf{B}'\mathbf{t}) = \left[2\exp\left((\mathbf{a} + \mathbf{B}\xi)'\mathbf{t} + \frac{1}{2}\mathbf{t}'(\mathbf{B}\Omega\mathbf{B}')\mathbf{t} \right) \Phi(\rho'\omega\mathbf{B}'\mathbf{t}) \right]$$

$$= \left[2\exp\left((\mathbf{a} + \mathbf{B}\xi)'\mathbf{t} + \frac{1}{2}\mathbf{t}'(\mathbf{B}\Omega\mathbf{B}')\mathbf{t} \right) \Phi(\tilde{\rho}'\tilde{\omega}\,\mathbf{t}) \right], \qquad (11)$$

$$\text{where } \tilde{\rho} = (\tilde{\omega})^{-1}\mathbf{B}\omega\,\rho$$

This proves the result. The two properties above provide the marginal distribution of the utilities under a MNP kernel mixed with skew normally distributed and normally distributed random coefficients, which is critical to the MACML estimation of the resulting model, as discussed next.

5.3 THE MODEL FRAMEWORK

We develop the model framework first in the context of a cross-sectional MNP model and then discuss the panel formulation. However, the skew-normal mixing can also be imposed on any other form of the MNP model, including settings with spatial dependencies and social dependencies across decision units, and combinations of temporal, spatial, and social dependencies.

5.3.1 Cross-Sectional MNP Formulation and Estimation

Consider a random-coefficients formulation in which the utility that an individual q $(q = 1, 2, ..., Q)$ associates with alternative i $(i = 1, 2, ..., I)$ is written as:

$$U_{qi} = \boldsymbol{\beta}_q' \mathbf{x}_{qi} + \boldsymbol{\gamma}_q' \mathbf{s}_{qi} + \tilde{\varepsilon}_{qi},\ \boldsymbol{\beta}_q \sim \text{MVSN}(\mathbf{b}, \boldsymbol{\omega}, \boldsymbol{\Omega}_+^*),$$
$$\boldsymbol{\Omega}_+^* = \begin{pmatrix} \mathbf{1} & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \boldsymbol{\Omega}^* \end{pmatrix}, \boldsymbol{\Omega}^* = (\boldsymbol{\omega})^{-1} \boldsymbol{\Omega} (\boldsymbol{\omega})^{-1}, \boldsymbol{\gamma}_q \sim \text{MVN}(\mathbf{c}, \boldsymbol{\Sigma}), \tilde{\varepsilon}_{qi} \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Psi}) \quad (12)$$

where $\mathbf{x}_{qi}$ is a $(D \times 1)$-column vector of exogenous attributes, $\mathbf{s}_{qi}$ is another $(K \times 1)$-column vector of exogenous attributes (including dummy variables for constants, except in one of the I alternative utilities), $\boldsymbol{\beta}_q$ is an individual-specific $(D \times 1)$-column vector of MVSN-distributed coefficients that varies across individuals based on unobserved individual attributes, $\boldsymbol{\gamma}_q$ is another individual-specific $(D \times 1)$-column vector of MVN-distributed coefficients that varies across individuals based on unobserved individual attributes (but with the coefficients on the dummy variables for the constants maintained as fixed coefficients in the vector $\boldsymbol{\gamma}_q$), and $\tilde{\boldsymbol{\varepsilon}}_q = (\tilde{\varepsilon}_{q1}, \tilde{\varepsilon}_{q2}, \tilde{\varepsilon}_{q3}, \ldots, \tilde{\varepsilon}_{qI})'$ is assumed to have a general covariance structure subject to identifiability considerations (let $\tilde{\boldsymbol{\varepsilon}}_\mathbf{q} \sim \text{MVN}_I(\mathbf{0}, \tilde{\boldsymbol{\Psi}})$). In many situations, such as in a path choice model (see Yai *et al.*, 1997) or a model with spatial location alternatives (see Bhat and Guo, 2007), a specific parametric structure, based on theoretical considerations appropriate to the context, can be placed on $\boldsymbol{\Psi}$. Similarly, in a pure random coefficients specification (as in Hausman and Wise, 1978), one may consider $\boldsymbol{\Psi}$ to be an identity matrix (or an identity matrix scaled by 0.5 or any other constant). Such specifications help in econometric identification as well as econometric efficiency. If a general covariance structure is adopted, there are many ways to ensure identification. An appealing approach is to take the differences of the error terms with respect to the first error term. Let $\varepsilon_{qi1} = (\tilde{\varepsilon}_{qi} - \tilde{\varepsilon}_{q1})$, and let $\boldsymbol{\varepsilon}_{\mathbf{q1}} = (\varepsilon_{q21}, \varepsilon_{q31}, ..., \varepsilon_{qI1})$. Then, up to a scaling factor, the covariance matrix of $\boldsymbol{\varepsilon}_{q1}$ (say $\boldsymbol{\Psi}_\mathbf{1}$) is identifiable. Next, scale the top left diagonal element of this error-

differenced covariance matrix to 1. Thus, there are $[(I-1)\times(I/2)]-1$ free covariance terms in the $(I-1)\times(I-1)$ matrix $\boldsymbol{\Psi}_1$. Finally, to ensure that whenever differences are taken with respect to the chosen alternative during the maximum approximate composite marginal likelihood (MACML) estimation, these differences are consistent with the same error covariance matrix $\tilde{\boldsymbol{\Psi}}$ for the undifferenced error term vector $\tilde{\boldsymbol{\varepsilon}}_q$, $\tilde{\boldsymbol{\Psi}}$ is constructed from $\boldsymbol{\Psi}_1$ by adding a top row of zeros and a first column of zeros (see Train, 2003; page 134). During the MACML estimation, then, we can obtain the $(I-1)\times(I-1)$ covariance matrix of the error differences taken with respect to the *m*th alternative as $\boldsymbol{\Psi}_{\mathbf{m}} = \boldsymbol{\Gamma}_{\mathbf{m}}\tilde{\boldsymbol{\Psi}}\boldsymbol{\Gamma}'_{\mathbf{m}}$, where $\boldsymbol{\Gamma}_{\mathbf{m}}$ is a $(I-1)\times I$ matrix which corresponds to the identity matrix of size $(I-1)$ with an extra column of –1's added as the *m*th column.

In Equation (12), we will assume that the random vectors $\boldsymbol{\beta}_q$, $\boldsymbol{\gamma}_q$, and $\tilde{\boldsymbol{\varepsilon}}_q$ are independent of each other for each individual, as well as that these vectors are independent of the corresponding coefficients of other individuals (this latter assumption can be relaxed within our modeling framework, as will be needed for accommodating spatial or social dependency effects). From the earlier definitions, we can write $\boldsymbol{\beta}_q = \mathbf{b} + \breve{\boldsymbol{\beta}}_q$ with $\breve{\boldsymbol{\beta}}_q \sim \text{MVSN}(\mathbf{0}, \boldsymbol{\omega}, \boldsymbol{\Omega}^*_+)$, and $\boldsymbol{\gamma}_q = \mathbf{c} + \breve{\boldsymbol{\gamma}}_q$ with $\breve{\boldsymbol{\gamma}}_q \sim \text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$. Also let $\mathbf{U}_q = (U_{q1}, U_{q2}, ..., U_{qI})'$ ($I\times 1$ vector), $\mathbf{x}_q = (\mathbf{x}_{q1}, \mathbf{x}_{q2}, ..., \mathbf{x}_{qI})'$ ($I \times D$ matrix), and $\mathbf{s}_q = (\mathbf{s}_{q1}, \mathbf{s}_{q2}, ..., \mathbf{s}_{qI})'$ ($I \times K$ matrix). Then, we can write:

$$\mathbf{U}_q = \left[\mathbf{x}_q\mathbf{b} + \mathbf{s}_q\mathbf{c}\right] + \left[\mathbf{x}_q\breve{\boldsymbol{\beta}}_q + \mathbf{s}_q\breve{\boldsymbol{\gamma}}_q + \tilde{\boldsymbol{\varepsilon}}_q\right], \tag{13}$$

Let $[.]_e$ indicate the e^{th} element of the column vector [.]. Equation (12) can equivalently be written using Equation (13) as:

$$U_{qi} = \left[\mathbf{x}_q\mathbf{b} + \mathbf{s}_q\mathbf{c}\right]_i + \left[\mathbf{x}_q\breve{\boldsymbol{\beta}}_q + \mathbf{s}_q\breve{\boldsymbol{\gamma}}_q + \tilde{\boldsymbol{\varepsilon}}_q\right]_i, \tag{14}$$

Define $V_{qi} = [\mathbf{x}_q\mathbf{b} + \mathbf{s}_q\mathbf{c}]_i$ and $\boldsymbol{\varepsilon}_{qi} = [\mathbf{x}_q\breve{\boldsymbol{\beta}}_q + \mathbf{s}_q\breve{\boldsymbol{\gamma}}_q + \tilde{\boldsymbol{\varepsilon}}_q]_i$. Also, assume that individual q chooses alternative m_q. In the utility differential form, we may write Equation (14) as:

$$u^*_{qim_q} = U_{qi} - U_{qm_q} = H_{qim_q} + \xi_{qim_q};\ H_{qim_q} = V_{qi} - V_{qm_q} \text{ and } \xi_{qim_q} = \varepsilon_{qi} - \varepsilon_{qm_q};\ i \neq m_q \qquad (15)$$

Then stack the utility differentials $u^*_{qim_q}\ (= U_{qi} - U_{qm_q}, i \neq m_q)$ in the following order: $\mathbf{u}^*_q = (u^*_{q1m_q}, u^*_{q2m_q}, ..., u^*_{qIm_q})'$, an $(I-1)\times 1$ vector. Correspondingly, let $\mathbf{H_q} = (H_{q1m_q}, H_{q2m_q}, ..., H_{qIm_q})'$, an $(I-1)\times 1$ vector, and define $\overline{\mathbf{\Omega}}_q = \mathbf{x}_q \mathbf{\Omega} \mathbf{x}'_q$ ($I \times I$ matrix), $\overline{\mathbf{\Sigma}}_q = \mathbf{s}_q \mathbf{\Sigma} \mathbf{s}'_q$ ($I \times I$ matrix) and $\mathbf{F}_q = \left[\overline{\mathbf{\Omega}}_q + \overline{\mathbf{\Sigma}}_q + \tilde{\mathbf{\Psi}}\right]$. Based on properties 1 and 2 earlier in the chapter, we can derive the location and other parameters of the vector $\mathbf{u}^*_q$, which is also skew-normally distributed. Specifically, by successive applications of property 2 and then property 1, we obtain the following important result:

$$\mathbf{u}^*_q \sim \text{MVSN}(\mathbf{H}_q, \breve{\boldsymbol{\omega}}_q, \breve{\mathbf{\Omega}}^*_{q+})], \qquad (16)$$

$$\breve{\mathbf{\Omega}}^*_{q+} = \begin{pmatrix} 1 & \breve{\boldsymbol{\rho}}'_q \\ \breve{\boldsymbol{\rho}}_q & \breve{\mathbf{\Omega}}^*_q \end{pmatrix},\ \breve{\mathbf{\Omega}}^*_q = (\breve{\boldsymbol{\omega}}_q)^{-1} \breve{\mathbf{\Omega}}_q (\breve{\boldsymbol{\omega}}_q)^{-1},\ \breve{\mathbf{\Omega}}_q = \mathbf{\Gamma}_{m_q} \mathbf{F}_q \mathbf{\Gamma}'_{m_q},$$
$$\breve{\boldsymbol{\rho}}_q = (\breve{\boldsymbol{\omega}}_q)^{-1} (\mathbf{\Gamma}_{m_q} \mathbf{x}_q) \boldsymbol{\omega} \boldsymbol{\rho}. \qquad (17)$$

$\breve{\boldsymbol{\omega}}_q$ is the diagonal matrix of standard deviations of $\breve{\mathbf{\Omega}}_q$. The parameters to be estimated include the **b** and **c** vectors, the elements of the covariance matrices $\mathbf{\Omega}$, $\mathbf{\Sigma}$, and $\tilde{\mathbf{\Psi}}$, and the $\boldsymbol{\rho}$ parameter vector. Collect all these elements into a single vector $\boldsymbol{\theta}$. Then, one can use the result above to obtain the likelihood contribution of individual *q* choosing alternative *m*, which takes the *I*-dimensional integral form below:

$$L_q(\boldsymbol{\theta}) = P(\mathbf{u}^*_q < 0) = \tilde{\Phi}_{I-1}\left((\breve{\boldsymbol{\omega}}_q)^{-1}(-\mathbf{H}_q); \breve{\mathbf{\Omega}}^*_{q+}\right) = 2\Phi_I\left(0, (\breve{\boldsymbol{\omega}}_q)^{-1}(-\mathbf{H}_q), \mathbf{\Omega}^*_{q-}\right). \qquad (18)$$

It is straightforward to see that if all the elements of $\boldsymbol{\rho}$ are zero, then the likelihood function above collapses to that of an MNP model. If not, the likelihood corresponds to a skew-normally mixed MNP model.

The *I*-dimensional integral in the likelihood contribution of each individual corresponds to the multivariate normal cumulative distribution function. The evaluation of such a function cannot be pursued using quadrature techniques due to the curse of dimensionality when the dimension of integration exceeds two (see Bhat, 2003). Consequently, the probability expression is typically approximated using Geweke-Hajivassiliou-Keane (GHK) simulator-based or the Genz-Bretz (GB) simulator-based techniques in the classical maximum simulated likelihood (MSL) inference approach (see Bhat *et al*., 2010b for a detailed description of these simulators) or using Markov Chain Monte Carlo (MCMC) techniques in the Bayesian inference approach (see Albert and Chib, 1993, McCulloch and Rossi, 2000, and Train, 2009). However, these MSL and Bayesian techniques can require extensive simulation, can be time-consuming, are not always very straightforward to implement, and can create convergence assessment problems as the number of dimensions of integration increases. On the other hand, the maximum approximate composite marginal likelihood (MACML) approach for estimation of MNP models, in which the MVNCD function is evaluated using an *analytic approximation* method, is quite accurate and very fast.

There is, however, one very important issue that still needs to be dealt with. This concerns the positive definiteness of several matrices in Equation (12). Specifically, for the estimation to work, we need to ensure the positive definiteness of the following matrices: $\boldsymbol{\Omega}^{*}_{+}$, $\boldsymbol{\Sigma}$, and $\tilde{\boldsymbol{\Psi}}$ (note that the positive definiteness of $\boldsymbol{\Omega}^{*}_{+}$ ensures the positive definiteness of $\boldsymbol{\Omega}^{*}$ and therefore $\boldsymbol{\Omega}$; this holds because of the property that any principal square sub-matrix of a positive definite matrix is also positive definite). Of these, one can guarantee the positive-definiteness of $\boldsymbol{\Sigma}$ and $\tilde{\boldsymbol{\Psi}}$ in a straightforward fashion using a Cholesky decomposition approach (by parameterizing the likelihood function in terms of the Cholesky-decomposed parameters). To guarantee the positive definiteness of the correlation matrix $\boldsymbol{\Omega}^{*}_{+}$, we use the approach of Bhat and Srinivasan (2005). Specifically, let $\mathbf{L}$ be the Cholesky decomposition matrix for $\boldsymbol{\Omega}^{*}_{+}$. We need to guarantee that the

parameters embedded within **L** are such that $\mathbf{\Omega}_{+}^{*}$ is a correlation matrix. This is done by parameterizing the diagonal terms of **L** as follows:

$$\mathbf{L}=\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{21} & \sqrt{1-l_{21}^2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{D+1,1} & l_{D+1,2} & l_{D+1,3} & \cdots & \sqrt{1-l_{D+1,1}^2-l_{D+1,2}^2\cdots-l_{D+1,D}^2} \end{bmatrix} \tag{19}$$

In the estimation, the Cholesky elements in the matrix **L** are estimated, guaranteeing that $\mathbf{\Omega}_{+}^{*}$ is indeed a correlation matrix.

5.3.2 Panel (or Repeated-Choice) MNP Formulation and Estimation

For the panel formulation, we introduce the index '*t*' for choice occasion. For ease in presentation, we will use the same number of choice occasions for each individual. Extension to the case of varying number of choice occasions per individual is straightforward.

Consider the random-coefficients formulation in which the utility that an individual q $(q=1,2,...,Q)$ associates at time period t $(t=1,2,...,T)$ with alternative i $(i=1,2,...,I)$ is written as:

$$\begin{aligned} &U_{qti}=\boldsymbol{\beta}_{q}'\mathbf{x}_{qti}+\boldsymbol{\gamma}_{q}'\mathbf{s}_{qti}+\tilde{\alpha}_{qi}+\tilde{\varepsilon}_{qti}, \\ &\boldsymbol{\beta}_{q}\sim \text{MVSN}(\mathbf{b},\boldsymbol{\omega},\mathbf{\Omega}_{+}^{*}),\ \mathbf{\Omega}_{+}^{*}=\begin{pmatrix}1 & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \mathbf{\Omega}^{*}\end{pmatrix},\ \mathbf{\Omega}^{*}=(\boldsymbol{\omega})^{-1}\mathbf{\Omega}(\boldsymbol{\omega})^{-1},\ \boldsymbol{\gamma}_{q}\sim \text{MVN}(\mathbf{c},\mathbf{\Sigma}). \end{aligned} \tag{20}$$

where all notations are as earlier except for the introduction of the index '*t*'. However, note that the vector $\mathbf{s}_{qti}$ is now a $(K\times 1)$-column vector of exogenous attributes without including a constant. $\tilde{\alpha}_{qi}$ is a normal random-effect term capturing time-stationary preference effects of individual *q* for alternative *i*. Also, as earlier, consider the (*I*×1)-vector $\tilde{\boldsymbol{\varepsilon}}_{qt}=(\tilde{\varepsilon}_{qt1},\tilde{\varepsilon}_{qt2},\tilde{\varepsilon}_{qt3},\ldots,\tilde{\varepsilon}_{qtI})'$, and assume that $\tilde{\boldsymbol{\varepsilon}}_{qt}\sim \text{MVN}(0,\tilde{\mathbf{\Psi}})$, with the same normalizations on $\tilde{\mathbf{\Psi}}$ as in the cross-sectional case (note that the $\tilde{\boldsymbol{\varepsilon}}_{qt}$ error terms are considered independent across individuals and choice occasions, and $\tilde{\boldsymbol{\varepsilon}}_{qt}$, $\boldsymbol{\beta}_{q}$, and $\boldsymbol{\gamma}_{q}$ are

also assumed independent for each individual *q*; $\boldsymbol{\beta}_q$ and $\boldsymbol{\gamma}_q$ are also independent across individuals). Next, stack the error terms $\tilde{\alpha}_{qi}$ into an (*I*×1)-vector $\tilde{\boldsymbol{\alpha}}_q = (\tilde{\alpha}_{q1}, \tilde{\alpha}_{q2}, \tilde{\alpha}_{q3}, \ldots, \tilde{\alpha}_{qI})'$ and let $\tilde{\boldsymbol{\alpha}}_q \sim \text{MVN}_I(\tilde{\mathbf{a}}, \tilde{\boldsymbol{\Lambda}})$. However, since only utility differentials matter, take the differentials of these random effects with respect to the first alternative $\alpha_{qi1} = \tilde{\alpha}_{qi} - \tilde{\alpha}_{q1}$. Then, only the mean vector $\mathbf{a} = [(\tilde{a}_2 - \tilde{a}_1), (\tilde{a}_3 - \tilde{a}_1, ..., (\tilde{a}_I - \tilde{a}_1)]$ and covariance matrix $\tilde{\boldsymbol{\Lambda}}_1$ of $\boldsymbol{\alpha}_{q1} = (\alpha_{q21}, \alpha_{q31}, ..., \alpha_{qI1})$ are identified. At the same time, whenever utility differences are taken with respect to the chosen alternative during the MACML estimation, these utility differences should be consistent with the same mean vector $\tilde{\mathbf{a}}$ and error covariance matrix $\tilde{\boldsymbol{\Lambda}}$ for the undifferenced error term vector $\tilde{\boldsymbol{\alpha}}_q$. To achieve this, we set $\tilde{a}_1 = 0$ (that is, the first element of the vector $\tilde{\mathbf{a}}$ is set to zero), and construct $\tilde{\boldsymbol{\Lambda}}$ from $\tilde{\boldsymbol{\Lambda}}_1$ by adding a first row of zeros and a first column of zeros.

We now set out some additional notation. Write $\tilde{\alpha}_{qi} = \tilde{a}_i + \breve{\alpha}_{qi}$, $\mathbf{A} = (\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_I)'$ ($I \times 1$ vector), $\breve{\boldsymbol{\alpha}}_q = (\breve{\alpha}_{q1}, \breve{\alpha}_{q2}, ..., \breve{\alpha}_{qI})'$ ($I \times 1$ vector) so that $\breve{\boldsymbol{\alpha}}_q \sim MVN_I(0, \tilde{\boldsymbol{\Lambda}})$. Define $\mathbf{U}_{qt} = (U_{qt1}, U_{qt2}, ..., U_{qtI})'$ ($I \times 1$ vector), $\mathbf{U}_q = (\mathbf{U}'_{q1}, \mathbf{U}'_{q2}, ..., \mathbf{U}'_{qT})'$ ($TI \times 1$ vector), $\tilde{\boldsymbol{\varepsilon}}_{qt} = (\tilde{\varepsilon}_{1t1}, \tilde{\varepsilon}_{qt2}, ..., \tilde{\varepsilon}_{qtI})'$ ($I \times 1$ vector), $\tilde{\boldsymbol{\varepsilon}}_q = (\tilde{\boldsymbol{\varepsilon}}'_{q1}, \tilde{\boldsymbol{\varepsilon}}'_{q2}, ..., \tilde{\boldsymbol{\varepsilon}}'_{qT})'$ ($TI \times 1$ vector), $\breve{\mathbf{A}}_q = \mathbf{1}_T \otimes \breve{\boldsymbol{\alpha}}_q$ ($TI \times 1$ vector), $\mathbf{x}_{qt} = (\mathbf{x}_{qt1}, \mathbf{x}_{qt2}, ..., \mathbf{x}_{qtI})'$ ($I \times D$ matrix), $\mathbf{x}_q = (\mathbf{x}'_{q1}, \mathbf{x}'_{q2}, ..., \mathbf{x}'_{qT})'$ ($TI \times D$ matrix), $\mathbf{s}_{qt} = (\mathbf{s}_{qt1}, \mathbf{s}_{qt2}, ..., \mathbf{s}_{qtI})'$ ($I \times K$ matrix), $\mathbf{s}_q = (\mathbf{s}'_{q1}, \mathbf{s}'_{q2}, ..., \mathbf{s}'_{qT})'$ ($TI \times K$ matrix). Let $\mathbf{1}_T$ be a column vector of ones of dimension *T*, and let $\mathbf{1}_{TT}$ be a matrix of ones of dimension *T*×*T*. Then, we can write:

$$\mathbf{U}_q = \left[\mathbf{x}_q \mathbf{b} + \mathbf{s}_q \mathbf{c} + (\mathbf{1}_T \otimes \mathbf{A})\right] + \left[\mathbf{x}_q \breve{\boldsymbol{\beta}}_q + \mathbf{s}_q \breve{\boldsymbol{\gamma}}_q + \breve{\mathbf{A}}_q + \tilde{\boldsymbol{\varepsilon}}_q\right]. \tag{21}$$

Let $[.]_e$ indicate the e^{th} element of the column vector [.], and let $d_{ti} = (t-1)I + i$. Equation (20) can be equivalently written using Equation (21) as:

$$U_{qti} = \left[\mathbf{x}_q\mathbf{b} + \mathbf{s}_q\mathbf{c} + (\mathbf{1}_T \otimes \mathbf{A})\right]_{d_{ti}} + \left[\mathbf{x}_q\breve{\boldsymbol{\beta}}_q + \mathbf{s}_q\breve{\boldsymbol{\gamma}}_q + \breve{\mathbf{A}}_q + \tilde{\boldsymbol{\varepsilon}}_q\right]_{d_{ti}}. \quad (22)$$

Define $V_{qti} = \left[\mathbf{x}_q\mathbf{b} + \mathbf{s}_q\mathbf{c} + (\mathbf{1}_T \otimes \mathbf{A})\right]_{d_{ti}}$ and $\boldsymbol{\varepsilon}_{qti} = \left[\mathbf{x}_q\breve{\boldsymbol{\beta}}_q + \mathbf{s}_q\breve{\boldsymbol{\gamma}}_q + \breve{\mathbf{A}}_q + \tilde{\boldsymbol{\varepsilon}}_q\right]_{d_{ti}}$. Also, assume that individual q chooses alternative m_{qt} at the t^{th} choice instance. In the utility differential form, we may write Equation (22) as:

$$\begin{aligned} &u^*_{qtim_{qt}} = U_{qti} - U_{qtm_{qt}} = H_{qtim_{qt}} + \xi_{qtim_{qt}};\ H_{qtim_{qt}} = V_{qti} - V_{qtm_{qt}} \text{ and} \\ &\xi_{qtim_{qt}} = \varepsilon_{qti} - \varepsilon_{qtm_{qt}};\ i \neq m_{qt} \end{aligned} \quad (23)$$

Then stack the utility differentials $u^*_{qtim_{qt}} (= U_{qti} - U_{qtm_{qt}}, i \neq m_{qt})$ in the following order: $\mathbf{u}^*_{qt} = (u^*_{qt1m_{qt}}, u^*_{qt2m_{qt}}, ..., u^*_{qtIm_{qt}})'$, an $(I-1)\times 1$ vector, and $\mathbf{u}^*_{\boldsymbol{q}} = \left[\left(\mathbf{u}^*_{\boldsymbol{q1}}\right)', \left(\mathbf{u}^*_{\boldsymbol{q2}}\right)', ..., \left(\mathbf{u}^*_{\boldsymbol{qT}}\right)'\right]'$, an $\left[(I-1)\times T\right]\times 1$ vector. Correspondingly, let $\mathbf{H}_{\boldsymbol{qt}} = (H_{qt1m_{qt}}, H_{qt2m_{qt}}, ..., H_{qtIm_{qt}})'$, an $(I-1)\times 1$ vector; $\mathbf{H}_{\boldsymbol{q}} = (\mathbf{H}'_{\boldsymbol{q1}}, \mathbf{H}'_{\boldsymbol{q2}}, ..., \mathbf{H}'_{\boldsymbol{qT}})'$, an $\left[(I-1)\times T\right]\times 1$ vector. It is easy to see that $\mathbf{u}^*_{\boldsymbol{q}}$ has a mean vector $\mathbf{H}_{\boldsymbol{q}}$. To determine the covariance matrix of $\mathbf{u}^*_{\boldsymbol{q}}$, a few additional matrix definitions are needed. Define $\overline{\boldsymbol{\Omega}}_q = \mathbf{x}_q\boldsymbol{\Omega}\mathbf{x}'_q$ ($TI\times TI$ matrix), $\overline{\boldsymbol{\Sigma}}_q = \mathbf{s}_q\boldsymbol{\Sigma}\mathbf{s}'_q$ ($TI\times TI$ matrix), $\overline{\boldsymbol{\Lambda}} = \left(\mathbf{1}_{TT} \otimes \tilde{\boldsymbol{\Lambda}}\right)$ ($TI\times TI$ matrix), and $\overline{\boldsymbol{\Psi}} = \mathbf{IDEN}_T \otimes \tilde{\boldsymbol{\Psi}}$ ($TI\times TI$ matrix). Let $\mathbf{F}_q = \left[\overline{\boldsymbol{\Omega}}_q + \overline{\boldsymbol{\Sigma}}_q + \overline{\boldsymbol{\Lambda}} + \overline{\boldsymbol{\Psi}}\right]$, and define $\boldsymbol{M}_{\boldsymbol{q}}$ as an $[(I-1)\times T]\times[TI]$ block diagonal matrix, with each block diagonal having $(I-1)$ rows and I columns corresponding to the q^{th} individual's t^{th} choice instance. This $(I-1)\times I$ matrix for individual q and observation time period t corresponds to an $(I-1)$ identity matrix with an extra column of -1's added as the $m_{qt}{}^{\text{th}}$ column. For instance, consider the case of $T = 2$, and $I = 4$. Let the qth individual be observed to choose alternative 2 in time period 1 and alternative 1 in time period 2. Then $\mathbf{M}_{\boldsymbol{q}}$ takes the form below.

$$\mathbf{M}_q = \left[\begin{array}{cccc|cccc} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \end{array}\right]. \quad (24)$$

Finally, we obtain the results below:

$$\mathbf{u}_q^* \sim \text{MVSN}(\mathbf{H}_q, \breve{\boldsymbol{\omega}}_q, \breve{\boldsymbol{\Omega}}_{q_+}^*)], \quad (25)$$

$$\breve{\boldsymbol{\Omega}}_{q_+}^* = \begin{pmatrix} 1 & \breve{\boldsymbol{\rho}}_q' \\ \breve{\boldsymbol{\rho}}_q & \breve{\boldsymbol{\Omega}}_q^* \end{pmatrix}, \breve{\boldsymbol{\Omega}}_q^* = (\breve{\boldsymbol{\omega}}_q)^{-1}\breve{\boldsymbol{\Omega}}_q(\breve{\boldsymbol{\omega}}_q)^{-1}, \breve{\boldsymbol{\Omega}}_q = \mathbf{M}_q\mathbf{F}_q\mathbf{M}_q', \ \breve{\boldsymbol{\rho}}_q = (\breve{\boldsymbol{\omega}}_q)^{-1}(\mathbf{M}_q\mathbf{x}_q)\boldsymbol{\omega}\boldsymbol{\rho}. \quad (26)$$

$\breve{\boldsymbol{\omega}}_q$ is the diagonal matrix of standard deviations of $\breve{\boldsymbol{\Omega}}_q$. The parameters to be estimated include the **A**, **b**, and **c** vectors, the elements of the covariance matrices $\boldsymbol{\Omega}$, $\boldsymbol{\Sigma}$, $\tilde{\boldsymbol{\Lambda}}$, and $\tilde{\boldsymbol{\Psi}}$, and the $\boldsymbol{\rho}$ parameter vector. Collect all these elements into a single vector $\boldsymbol{\theta}$. Then, one can use the result above to obtain the likelihood contribution of individual *q* choosing alternative *m*, which takes the $[T(I-1)+1]$-dimensional integral form below:

$$L_q(\boldsymbol{\theta}) = P(\mathbf{u}_q^* < 0) = \tilde{\Phi}_{T(I-1)}\left((\breve{\boldsymbol{\omega}}_q)^{-1}(-\mathbf{H}_q); \breve{\boldsymbol{\Omega}}_{q_+}^*\right) = 2\Phi_{T(I-1)+1}\left(\mathbf{0}, (\breve{\boldsymbol{\omega}}_q)^{-1}(-\mathbf{H}_q), \breve{\boldsymbol{\Omega}}_{q-}^*\right). \quad (27)$$

In this panel setting, the parameter vector $\boldsymbol{\theta}$ is estimated by defining "events" in the MACML procedure as the pairs of choice observations across the choice occasions of the individual. Letting the individual's choice at time *t* be denoted by the index C_{qt}, the CML function for individual *q* is:

$$\begin{aligned} L_{CML,q}(\boldsymbol{\theta}) &= \prod_{t=1}^{T-1}\prod_{w=t+1}^{T} \text{Prob}\,(C_{qt} = m_{qt}, C_{qw} = m_{qw}) \\ &= \prod_{t=1}^{T-1}\prod_{w=t+1}^{T} \text{Prob}[\mathbf{u}_{qt}^* < 0 \text{ and } \mathbf{u}_{qw}^* < 0] \\ &= \prod_{t=1}^{T-1}\prod_{w=t+1}^{T} \text{Prob}[\tilde{\mathbf{u}}_{qtw}^* < 0] \end{aligned} \quad (28)$$

where $\ddot{\mathbf{u}}^{*}_{qtw} = \left[\left(\mathbf{u}^{*}_{qt}\right)', \left(\mathbf{u}^{*}_{qw}\right)'\right]'$. The computational effort is reduced in the CML above because only pairwise marginal multivariate probabilities are being considered across choice occasions. However, each multivariate orthant probability above still has a dimension equal to $[(I-1)\times 2]+1$:

$$P(\ddot{\mathbf{u}}^{*}_{qtw} < 0) = \widetilde{\Phi}_{2\times(I-1)}\left((\ddot{\boldsymbol{\omega}}_{qtw})^{-1}(-\ddot{\mathbf{H}}_{qtw}); \ddot{\boldsymbol{\Omega}}^{*}_{qtw+}\right) = 2\Phi_{2\times(I-1)+1}\left(0, (\ddot{\boldsymbol{\omega}}_{qtw})^{-1}(-\ddot{\mathbf{H}}_{qtw}), \ddot{\boldsymbol{\Omega}}^{*}_{qtw-}\right), \quad (29)$$

where $\ddot{\mathbf{H}}_{qtw} = (\mathbf{H}'_{qt}, \mathbf{H}'_{qw})'$, $\ddot{\boldsymbol{\Omega}}^{*}_{qtw+}$ and $\ddot{\boldsymbol{\Omega}}^{*}_{qtw-}$ are appropriate sub-matrices of $\breve{\boldsymbol{\Omega}}^{*}_{q+}$ and $\breve{\boldsymbol{\Omega}}^{*}_{q-}$, respectively (that is, they include elements corresponding to the t^{th} and w^{th} choice occasions of the individual). But such an orthant probability is conveniently computed using the approximation part of the MACML, leading to solely bivariate and univariate cumulative normals.

5.4 SIMULATION ANALYSIS

In this section, we undertake a simulation experiment with two objectives in mind. The first objective is to examine the ability of the MACML estimation method to recover parameters in the MNP model with skew-normally distributed coefficients. The second objective is to illustrate the problems that may arise from ignoring the skewness in the random coefficient distribution, which is equivalent to assuming that the distribution is normally distributed when it actually is not.

5.4.1 Experimental Set-Up

A cross-sectional formulation is used for the simulation experiments. Two cases are considered: (1) a three alternative case with three exogenous variables and (2) a five alternative case with five exogenous variables. In both the cases, the values of each of the exogenous variables for the alternatives are drawn from a standard univariate normal distribution. In particular, a sample of 5000 realizations of the exogenous variables is generated corresponding to 5000 individuals. The first case specifies a skew-normally distributed random coefficient vector $\boldsymbol{\beta}_q$ on all the three exogenous variables, and the second case specifies a skew-normally distributed random coefficient vector $\boldsymbol{\beta}_q$ on the

first three exogenous variables and a normally distributed random coefficient vector $\boldsymbol{\gamma}_q$ on the remaining two exogenous variables. For the five-dimensional simulation case, the coefficient vector $\boldsymbol{\gamma}_q$ is assumed to be a realization from $\boldsymbol{\gamma}_q \sim \text{MVN}(\mathbf{c}, \boldsymbol{\Sigma})$, with:

$$\mathbf{c} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}. \tag{30}$$

In the simulation experiments, the coefficient vector $\boldsymbol{\beta}_q$ is assumed to be a realization from $\boldsymbol{\beta}_q \sim \text{MVSN}(\mathbf{b}, \boldsymbol{\omega}, \boldsymbol{\Omega}^*_+)$, $\boldsymbol{\Omega}^*_+ = \begin{pmatrix} 1 & \boldsymbol{\rho}' \\ \boldsymbol{\rho} & \boldsymbol{\Omega}^* \end{pmatrix}$, $\boldsymbol{\Omega}^* = (\boldsymbol{\omega})^{-1}\boldsymbol{\Omega}(\boldsymbol{\omega})^{-1}$, where

$$\mathbf{b} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}, \ \boldsymbol{\omega} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1.25 \end{pmatrix}, \text{ and } \boldsymbol{\Omega}^*_+ = \begin{pmatrix} 1 & -0.7 & -0.7 & -0.7 \\ -0.7 & 1 & 0.49 & 0.49 \\ -0.7 & 0.49 & 1 & 0.49 \\ -0.7 & 0.49 & 0.49 & 1 \end{pmatrix}. \tag{31}$$

The correlation matrix $\boldsymbol{\Omega}^*_+$ above is constructed in a specific manner so that the off-diagonal elements of the corresponding Cholesky matrix are all zero, except for the first column which now contains the skew parameters (= –0.7) as its elements.[21] Essentially, this way of constructing the correlation matrix assumes that all the correlations in the augmented four-dimensional correlation matrix (corresponding to the three-dimensional skew-normally distributed random coefficient vector) originates in the skew distribution of the coefficients, with no residual correlation beyond that generated by the skew. Such a specification is parsimonious, and can be used to reduce the number of parameters to be estimated in the skew-normal MNP model. For instance, in the MNP with three skew-normal coefficients, there is a reduction from nine correlation parameters to just three. More generally, in a model with D skew-normal coefficients, there is a reduction from

[21] The Cholesky matrix of $\boldsymbol{\Omega}^*_+$ is $\mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -0.7 & 0.7141 & 0 & 0 \\ -0.7 & 0 & 0.7141 & 0 \\ -0.7 & 0 & 0 & 0.7141 \end{pmatrix}$

$\frac{D(D+1)}{2}+D$ to D parameters in the augmented correlation matrix. Clearly, this can be an effective way to allow a large number of skew-normally distributed coefficients without an explosion in the number of model parameters to be estimated. The other benefit of such a specification is that the skew parameter vector $\boldsymbol{\rho}$ is directly estimated because it "sits" as the first column of the Cholesky matrix (minus the first row element).

Another point to note about our skew specification for the $\boldsymbol{\beta}_q$ vector is that the negative values for $\mathbf{b}$ and $\boldsymbol{\rho}$ provide a negative location parameter and leftward skew for the marginal distributions of each of the $\boldsymbol{\beta}_q$ coefficients that is similar to a (negative) log-normal distribution. Such a specification may be considered for cost and other coefficients. Of course, in reality, the skew-normal distribution can be used for all parameters to allow a range of "seamless" and "continuous" marginal distribution possibilities that ranges from normality to non-normality.

The method to generate realizations from the MVSN distribution for $\boldsymbol{\beta}_q$ is based on first drawing a multivariate standard normal vector with correlation matrix $\boldsymbol{\Omega}_+^*$ in the usual way. This constitutes a draw for the latent underlying $(D+1)$-variate normally distributed vector $\tilde{\mathbf{M}}=(\tilde{M}_1,\tilde{\mathbf{M}}_2')'$, where $\tilde{M}_1$ is a latent (1×1)-vector and $\tilde{\mathbf{M}}_2$ is a $(D\times1)$-vector (see Equation (5); D = 3 in the current case). From this multivariate standard normal draw, a D-variate vector from the multivariate standard skew normal distribution is generated as follows:

$$\mathbf{Z}=\begin{cases}\tilde{\mathbf{M}}_2 & \text{if } \tilde{M}_1>0\\ -\tilde{\mathbf{M}}_2 & \text{if } \tilde{M}_1\le 0.\end{cases} \tag{32}$$

Finally, the error term vector $\boldsymbol{\varepsilon}_q=(\varepsilon_{q1},\varepsilon_{q2},\varepsilon_{q3},\ldots,\varepsilon_{qI})'$ is drawn from $\boldsymbol{\varepsilon}_q \sim \text{MVN}(\mathbf{0}, 0.5\times\mathbf{IDEN}_I)$, where $\mathbf{IDEN}_I$ is the identity matrix of dimension I (in the notation of Equation (12), $\boldsymbol{\Psi}=0.5\times\text{IDEN}_I$). Thus, we assume and maintain the IID

normal assumption for $\boldsymbol{\varepsilon}_q$ in the current simulation experiment. The alternative with the highest utility for each individual q is then identified as the chosen alternative.

The above data generation process is undertaken 40 times with different realizations of the $\boldsymbol{\beta}_q, \gamma_q$, and $\boldsymbol{\varepsilon}_q$ vectors $(q = 1, 2, ..., Q)$ to generate 40 different data sets. The MACML estimator is applied ten times to each dataset, with different sets of permutations (across the ten runs on the same dataset) to decompose the multivariate normal cumulative distribution or MVNCD function into a product of marginal and conditional probabilities (see Bhat, 2011). In each of the ten runs on the same dataset, ten different random permutations are generated and used for each individual (the random permutation varies across individuals) to approximate the MVNCD function for that individual. The approximation error for each parameter (due to using the analytic approximation to the MVNCD function) is obtained by computing the standard deviation of estimated parameters among the 10 different parameter estimates on the same data set.

A number of performance measures are identified to assess the performance of the MACML approach in being able to recover the underlying "true" parameters (which is the first objective of our simulation exercise). The performance measures, and the various steps to compute these measures, are described below:

(1) Estimate the MACML parameters for each data set *s* and for each of 10 independent sets of permutations for computing the MVNCD function.

(2) For each data set *s*, estimate the standard errors (s.e.) (using the sandwich covariance matrix estimator; see McFadden and Train, 2000).

(3) For each data set *s*, compute the mean estimate for each model parameter across the 10 random permutations used. Label this as MED, and then take the mean of the MED values across the data sets to obtain **a mean estimate**. Compute the **absolute percentage bias** (APB) as:

$$APB = \left| \frac{\text{mean estimate - true value}}{\text{true value}} \right| \times 100 .$$

(4) For each data set *s*, compute the median s.e. for each model parameter across the 10 draws. Call this MSED, and then take the mean of the MSED values across the 40 data sets and label this as **the asymptotic standard error.**

(5) Next, compute the standard deviation of the MED values across the 40 data sets to obtain the finite sample standard error for each parameter, and label this as the **empirical standard error.** Note that the asymptotic standard error is essentially an approximation to this empirical standard error, and the consistency of the estimator for the asymptotic standard error implies that the asymptotic and empirical standard error estimates should be close to one another.

(6) Next, for each data set *s*, compute the approximation standard deviation for each parameter as the standard deviation in the estimated parameter values across the 10 independent permutations (about the MED value). Call this standard deviation as APPMED. For each parameter, take the mean of APPMED across the different data sets. Label this as the **approximation standard error** for each parameter.

(7) For each parameter, compute an **approximation adjusted asymptotic standard error** as follows: $\sqrt{(\text{asymptotic standard error})^2 + (\text{approximation standard error})^2}$. Similarly, compute an **approximation adjusted empirical standard error** as follows: $\sqrt{(\text{empirical standard error})^2 + (\text{approximation standard error})^2}$.

The second objective of examining the implications of ignoring skewness when actually present is achieved by generating data exactly as discussed above. Once generated, we estimate a simple normally-mixed MNP model on the data, assuming (incorrectly) that $\boldsymbol{\rho} = 0$ (using ten random permutation per individual in the computation of the MVNCD function, exactly as earlier). We will refer to this model as the MNP-normal (or MNP-N) model. We compare this MNP-N model with the skew normally-mixed (or MNP-SN) model. For this comparison, we ignore approximation error issues and undertake a single MNP-N estimation on each of the 40 datasets generated. We then randomly pick one of the MNP-SN model estimates for each of the 40 datasets (as already estimated earlier), and use that to compare with the MNP-N model. The performances of the two models are evaluated by (1) comparing the mean APB values

across parameters and (2) undertaking a likelihood ratio test (LRT) for each of the 40 datasets. For the mean APB computation, the APB in the skewness parameters is not included in the MNP-SN model because the MNP-N implicitly assumes that $\boldsymbol{\rho} = 0$ (this allows an "apples to apples" comparison between the MNP-N and MNP-SN models). For the likelihood ratio test, we compare the test statistic for each data set with the table chi-squared distribution value with three degrees of freedom (corresponding to each of the three skew parameters in the $\boldsymbol{\rho}$ vector being zero). The number of times out of the 40 data sets that the MNP-SN model rejects the MNP-N model is then obtained, along with the mean value of the LRT statistic across the 40 data sets.

5.4.2 Simulation Results

5.4.2.1 Ability of MACML to Recover Model Parameters

The results for the first objective of evaluating the ability to recover model parameters are summarized in the Table 5.1 for the three alternative case with three exogenous variables, and in Table 5.2 for the five alternative case with five exogenous variables.

5.4.2.1.1 The Three Alternative Case with Three Exogenous Variables

The results in Table 5.1 for the three alternative case indicate that the MACML method does reasonably well in recovering the true parameters. The absolute percentage bias (APB) ranges from 7.1% to 11.2% across the parameters, with a mean value of 9.2% (see the last row of the table under the "absolute percentage bias" column). The APB values are generally somewhat smaller and more stable (across parameters) for the location parameters of the distributions of the $\boldsymbol{\beta}_q$ parameter vector (*i.e.*, the *b* parameter estimates in the table) than for the skew parameter estimates (*i.e.*, the ρ values) or the scale parameter estimates of the distribution of the $\boldsymbol{\beta}_q$ parameter vector (*i.e.*, the ω parameters in the table). This is not surprising, because the *b* parameters enter more linearly in the likelihood function of Equation (18) (through the mean of the MVNCD function) than do the skew and scale parameters (that enter more non-linearly and in a complex manner through the covariance matrix of the MVNCD function).

Table 5.1. Simulation Results for the Three Alternative-Three Variable Case

Parameter	True Value	Parameter Estimates		Standard Error (SE) Estimates				
		Mean Estimate	Absolute Percentage Bias	Asymptotic SE	Empirical SE	Approximation SE	Approximation Adjusted Asymptotic SE	Approximation Adjusted Empirical SE
Location parameters of the β_q vector								
$b1$	-1.000	-0.906	9.4%	0.116	0.134	0.073	0.137	0.153
$b2$	-1.000	-0.917	8.3%	0.114	0.125	0.072	0.135	0.144
$b3$	-1.000	-0.932	6.8%	0.122	0.127	0.076	0.144	0.149
Skewness parameters of the β_q vector								
$\rho 1$	-0.700	-0.770	10.1%	0.065	0.081	0.048	0.081	0.094
$\rho 2$	-0.700	-0.778	11.2%	0.062	0.064	0.047	0.078	0.079
$\rho 3$	-0.700	-0.750	7.1%	0.061	0.070	0.044	0.076	0.083
Scale parameters of the β_q vector								
$\omega 1$	1.000	1.112	11.2%	0.135	0.144	0.073	0.154	0.162
$\omega 2$	1.000	1.111	11.1%	0.134	0.122	0.068	0.150	0.140
$\omega 3$	1.250	1.344	7.5%	0.150	0.135	0.080	0.170	0.157
Overall Mean Value			9.2%	0.107	0.111	0.065	0.125	0.129

One can also observe that all the parameters associated with the third variable are recovered better than the first two variables, perhaps because of the higher standard deviation of this coefficient (=1.25) relative to the other two coefficients. When there is higher variation in a coefficient, it provides more information in the data to pin down the moments of its distribution.

The asymptotic and empirical standard error values (reflecting sampling standard error) are quite close to one another, reflecting the consistency of the MACML estimator of the asymptotic covariance matrix. These sampling standard error estimates of the parameters indicate good efficiency of the MACML estimator, with the standard errors being between 8%-15% of the mean values of the estimator. Also, the approximation standard error estimates are smaller than the sampling standard errors. On average, the approximation standard error is about 60% of the corresponding asymptotic and empirical standard error values. On the other hand, in a similar simulation setting, the approximation standard error of the MACML estimator with just one permutation per individual (as opposed to ten used here) was found to be only of the order of 13% of the sampling standard errors when the MACML approach was applied to a strictly normally-distributed coefficients model (see Chapter 2). Clearly, even though the skew-normally distributed coefficients can be viewed as originating from an augmented and truncated multivariate normal distribution, and the cumulative distribution function of the skew-normal distribution may be written as that of a normal distribution function with an added dimension, the introduction of asymmetry does appear to introduce more approximation error in the MACML approach. This is an issue that needs further examination in the future. Nonetheless, this should not detract from the fact that the MACML estimator still does very well. In fact, the final column provides the approximation-adjusted asymptotic and empirical standard errors for the MACML estimator, which are only 13-25% higher than the corresponding unadjusted standard errors. Also, the approximation-adjusted standard errors are still only 10-17% of the corresponding mean values of the estimators, indicating that the approximation standard errors introduced by the MACML approach are small in the larger inference context.

5.4.2.1.2 Five Alternative Case with Five Variables

The results for the five alternative case with five variables are summarized in Table 5.2. The APB is of the same order as that in the case with three skew-normal coefficients, and ranges from 3% to 18.5% with a mean of 9.4%. As in the previous section, the APB values are smaller and more stable for the *b* parameter estimates than for the ρ and ω parameter estimates. Further, there is a clear increase in the APB values for the ρ and ω parameter estimates compared to the case with three coefficients. However, the APB for the parameters characterizing the normally distributed coefficients (see the *c* and the σ parameters in the fourth and fifth row panels of Table 5.2, respectively) are estimated very well, with the APBs ranging from 3-6.5% (mean of the APBs for these parameters is 4.5%, which is less than half of the overall mean APB of 9.4%).

The sampling (asymptotic and empirical) standard error values of the parameters continue to indicate good efficiency of the MACML estimator, with the sampling standard errors ranging between 5%-14% of the mean values of the estimator. Also, the approximation standard error estimates continue to be smaller than the sampling standard error estimates. On average, the approximation standard errors are about 45% of the corresponding asymptotic standard error estimates and 40% of the corresponding empirical standard error values, which is even better than the three-dimensional case. While the approximation errors are close to the sampling standard errors for the skewness elements ρ, this is because the standard errors are extremely small for these elements in the first place. At the end, the approximation-adjusted asymptotic and empirical standard errors are only 5-16% of the mean values of the estimator, which is about the same range as the unadjusted standard errors as a percentage of the mean values.

To summarize, the MACML inference approach does very well in recovering the parameters in a skew-normally mixed MNP model (with or without normally mixed coefficients). However, there is also evidence that there is some kind of a relative degradation of performance when skew-normally distributed coefficients are introduced (relative to the case when there are only normally-distributed coefficients, in which case the MACML approach does extremely well).

Table 5.2. Simulation Results for the Five Alternative-Five Variable Case

Parameter	True Value	Parameter Estimates		Standard Error (SE) Estimates				
		Mean Estimate	Absolute Percentage Bias	Asymptotic SE	Empirical SE	Approx. SE	Approx. Adjusted Asymptotic SE	Approx. Adjusted Empirical SE
Location parameters of the β_q vector								
$b1$	-1.000	-0.914	8.6%	0.107	0.120	0.053	0.119	0.132
$b2$	-1.000	-0.917	8.3%	0.106	0.137	0.053	0.119	0.147
$b3$	-1.000	-0.990	1.0%	0.116	0.135	0.058	0.130	0.147
Skewness parameters of the β_q vector								
$\rho 1$	-0.700	-0.825	17.9%	0.036	0.042	0.030	0.047	0.051
$\rho 2$	-0.700	-0.824	17.7%	0.036	0.044	0.028	0.046	0.052
$\rho 3$	-0.700	-0.769	9.9%	0.034	0.036	0.030	0.046	0.046
Scale parameters of the β_q vector								
$\omega 1$	1.000	1.184	18.4%	0.144	0.167	0.067	0.159	0.180
$\omega 2$	1.000	1.168	16.8%	0.143	0.152	0.066	0.157	0.166
$\omega 3$	1.250	1.381	10.5%	0.158	0.162	0.067	0.172	0.175
Mean values of the γ_q vector								
$c1$	1.000	1.041	4.1%	0.107	0.107	0.038	0.114	0.114
$c2$	1.000	1.039	3.9%	0.107	0.112	0.038	0.113	0.118
Covariance elements of the γ_q vector								
$\sigma 1$	1.000	1.065	6.5%	0.126	0.144	0.044	0.134	0.151
$\Sigma 12$	0.500	0.516	3.2%	0.067	0.059	0.022	0.071	0.063
$\sigma 2$	1.000	1.051	5.1%	0.124	0.142	0.045	0.132	0.149
Overall Mean Value			9.4%	0.101	0.111	0.046	0.111	0.121

Some of this degradation is surely attributable to the more difficult asymmetric shapes that need to be characterized with skew-normal distributions. More explorations are needed to examine such behavior. However, despite the relative degradation, the MACML model is able to recover all parameters well, with the approximation errors being quite inconsequential in the larger sampling inference context.

5.4.2.2 Effects of Ignoring Skewness in the Coefficient Distribution

This section focuses on the implications of ignoring skewness when actually present. The results are presented in Table 5.3 for both the three dimensional case (three alternatives-three variable case) and the five dimensional case (five alternatives-five variable case). The results clearly show the poor performance of the MNP-N model (which assumes away any skewness) relative to the MNP-SN model (which explicitly accommodates skewness). The mean APB value across the location parameters is of the order of 60% in the MNP-N model compared to the corresponding mean APB value of 6-8% from the MNP-SN model. The scale parameters also have a larger mean APB in the MNP-N model compared to the MNP-SN model. Overall, the use of a normal distribution when there is skew in the random parameters can lead to seriously mis-estimated distributions for the random parameters. This, in turn, will then lead to mis-estimated willingness to pay and welfare measures. An interesting observation from the five-dimensional analysis, though, is that if there are truly normally distributed coefficients in the model, these do not appear to be substantially affected by mis-specifications on the other coefficients (as can be noticed from the similar mean APB values for the mean elements of the γ_q vector and the covariance elements of the γ_q vector). The log-likelihood values at convergence from the MNP-SN model is always better than from the MNP-N model in all the 40 generated data sets. The mean value of the log-likelihood ratio statistic across all the 40 data sets for each of the three-dimensional and five-dimensional cases is provided in Table 5.3. Also, for each and every data set, the log-likelihood ratio statistic is higher than the corresponding chi-squared table value (see the last row of Table 5.3).

Table 5.3. Effects of Ignoring Skewness in the Mixing Distribution (when present)

Evaluation Metric	Three Dimensional Case		Five Dimensional Case	
	Skew Normal	Normal	Skew Normal	Normal
Mean APB				
Location parameters of the $\boldsymbol{\beta}_q$ vector	7.7%	58.8%	5.8%	60.4%
Scale parameters of the $\boldsymbol{\beta}_q$ vector	8.8%	18.3%	15.4%	18.3%
Mean values of the $\boldsymbol{\gamma}_q$ vector	-	-	4.1%	3.4%
Covariance elements of the $\boldsymbol{\gamma}_q$ vector	-	-	5.1%	4.3%
Across all parameters $\boldsymbol{\beta}_q$ and $\boldsymbol{\gamma}_q$ vector	8.3%	38.6%	7.9%	23.3%
Mean log- likelihood value at convergence	-2056.6	-2095.0	-4132.3	-4219.7
Mean value of the log-likelihood ratio statistic across datasets	76.9		174.8	
Number of times the likelihood ratio test (LRT) favors the skew normal model	Every Time when compared to $\chi^2_3 = 11.34$		Every Time when compared to $\chi^2_3 = 11.34$	

Overall, the results clearly highlight the bias in characterizing the distribution of random coefficients if skewness effects in the coefficients are ignored when actually present.

5.5 CONCLUSION

In this chapter, we propose the use of the multivariate skew-normal distribution function to accommodate non-normal mixing in MNP models. The multivariate skew normal (MSN) distribution retains several attractive properties of the multivariate normal distribution. It is tractable, parsimonious in parameters that regulate the distribution and its skewness, and includes the normal distribution as a special interior point case. It also is a very flexible unimodal density structure that allows a "seamless" and "continuous" variation from normality to non-normality, and can replicate a variety of smooth unimodal density shapes. At the same time, we propose the use of an MNP kernel because the combination of skew-normal mixing over the MNP kernel lends itself

perfectly to estimation using the maximum approximate composite marginal likelihood (MACML) approach. This is because of two properties of the skew distribution. The first is that it is closed under any affine transformation of the skew-normally distributed vector, and the second is that it is closed under the sum of a skew-normally distributed vector and a normally distributed vector of the same dimensions. As importantly, the cumulative distribution function of the *D*-variate skew normally distributed variable requires only the evaluation of a $(D+1)$-dimensional multivariate cumulative normal distribution function. All of these properties are gainfully exploited in the study to formulate an MNP model with non-normal mixing, while also being able to estimate the model in a simple and computationally efficient MACML approach. To our knowledge, this is the first study to propose and formulate a skew-normally mixed MNP model.

A simulation exercise is undertaken to evaluate the ability of the proposed approach to recover parameters in the skew-normally mixed MNP model. Two cases are considered: (1) a three alternative case with three exogenous variables and (2) a five alternative case with five exogenous variables. The first case considers a three-variate skew normal distribution for the coefficients on the three exogenous variables, while the second case considers a three-variate skew normal distribution for three variables and a bivariate normal for two variables. The results show that our proposed approach does very well in recovering the parameters in a skew-normally mixed MNP model. In addition, the simulation results clearly highlight the bias in characterizing the distribution of random coefficients as well as the poor data fit if skewness, when actually present, is ignored away.

CHAPTER 6: A MORE ACCURATE MACML ESTIMATION USING THE SECOND ORDER APPROXIMATION OF THE MVNCD FUNCTION

6.1 INTRODUCTION

The simulation studies performed in chapter 2 for the different model structures indicate that the MACML approach performs extremely well in recovering parameters in the case of a normally distributed random coefficient model, however, the parameter recovery for the skew-normally distributed random coefficient models of chapter 5 were not as good as the ones observed for the normally distributed random coefficient models. In this chapter we will consider the case for improving the MACML approach by extending it to the second order MACML method which incorporates the second order MVNCD approximation proposed in Joe (1995). We will briefly review the first order MVNCD function used in the MACML method and then consider the second order MACML method which incorporates the second order MVNCD approximation proposed in Joe (1995).

6.1.1 Multivariate Standard Normal Cumulative Distribution Function

As described in the introduction chapter of this dissertation, the approximation used in the MACML method of Bhat (2011a) can be described by considering a normal random vector $\boldsymbol{W}$ of dimension I $(W_1, W_2, W_3, \ldots, W_I)$ with mean zero and variance 1 for each of the dimensions. The correlation matrix is assumed to be $\boldsymbol{\Sigma}$. Now, the orthant probability that we intend to evaluate can be written as:

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1,\ W_2 < w_2,\ W_3 < w_3, ...,\ W_I < w_I) \tag{1}$$

This can be written as a product of conditional probabilities as given below for $I \geq 3$:

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1,\ W_2 < w_2) \times \prod_{i=3}^{I} \Pr(W_i < w_i \mid W_1 < w_1,\ W_2 < w_2,\ W_3 < w_3, ...,\ W_{i-1} < w_{i-1}). \tag{2}$$

Next, we define the indicator variable $\tilde{I}_i = I(W_i < w_i)$, where $I(\cdot)$ is the indicator function that takes a values 1 if the condition within the parenthesis is true and zero otherwise. This also implies that $E(\tilde{I}_i) = \Phi(w_i)$, where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a univariate normal distribution. Using these notations, equation (2) can be written as:

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1,\ W_2 < w_2) \times \prod_{i=3}^{I} E(\tilde{I}_i \mid \tilde{I}_1 = 1,\ \tilde{I}_2 = 1, \tilde{I}_3 = 1, ...,\ \tilde{I}_{i-1} = 1). \qquad (3)$$

In equation (3) above, the first term is computed using a bivariate CDF and the following approximation is used for the conditional probabilities:

$$\begin{aligned} E(\tilde{I}_i \mid \tilde{I}_1 = 1,\ \tilde{I}_2 = 1, \tilde{I}_3 = 1, ...,\ \tilde{I}_{i-1} = 1) & \\ = E(I_i) + (\boldsymbol{\Omega}_{<i}^{-1} \cdot \boldsymbol{\Omega}_{i,<i})'(1 - E(I_1),\ 1 - E(I_2) ... 1 - E(I_{i-1}))' & \qquad (4) \\ = \Phi(w_i) + (\boldsymbol{\Omega}_{<i}^{-1} \cdot \boldsymbol{\Omega}_{i,<i})'(1 - \Phi(w_1),\ 1 - \Phi(w_2) ... 1 - \Phi(w_{i-1}))' & \end{aligned}$$

Where, $\boldsymbol{\Omega}_{i,<i}$ is a column vector formed from the covariance elements $\mathrm{Cov}(\tilde{I}_i, \tilde{I}_j)$ where j =1, 2….i–1 and $\boldsymbol{\Omega}_{<i}$ is an (i–1) × (i–1) matrix whose (j, k) element is formed by covariance elements $\mathrm{Cov}(\tilde{I}_j, \tilde{I}_k)$ with j and k satisfying the following condition: 1≤ j, k ≤ i–1. The covariance term can be evaluated as $\mathrm{Cov}(\tilde{I}_j, \tilde{I}_k) = E(\tilde{I}_j \tilde{I}_k) - E(\tilde{I}_j)E(\tilde{I}_k)$. An additional note to be made about the first order MVNCD approximation is that the initial decomposition of the function into a bivariate CDF and the product of conditional probabilities can be done in $I!/2$ permutations. Each of these decompositions yields a different probability and Joe (1995) proposes averaging over all the permutations to obtain an estimate. Bhat (2011a) instead proposes the use of only 1 random permutation for each observation and as the results of the simulation experiments in the second chapter of this dissertation shows, good estimation properties are observed with using just one permutation.

6.1.2 Second Order Extension of the MVNCD

Joe (1995) extends the MVNCD method presented above further by proposing an improved second-order approximation. This second-order approximation uses trivariate and four-variate CDF functions, in addition to the univariate and bivariate CDFs to approximate the conditional probability of equation (2) and equation (3). The idea behind the extension is to include second-order interaction terms of the type $\tilde{I}_{i'j'} = \tilde{I}_{i'}\tilde{I}_{j'}$ also in the conditioning and hence in the approximation of equation (4). This can be written as below:

$$\begin{aligned} &E(\tilde{I}_i \mid \tilde{I}_1 = 1,\ \tilde{I}_2 = 1, \tilde{I}_3 = 1, ...,\ \tilde{I}_{i-1} = 1; \tilde{I}_{jk} = 1, 1 \le j,k \le i-1) = E(I_i) + (\mathbf{\Omega}^{*}_{<i}{}^{-1} \cdot \mathbf{\Omega}^{*}_{\boldsymbol{i},<\boldsymbol{i}})' \\ &(1-E(I_1),\ ...1-E(I_{i-1}), 1-E(I_{12}), ...1-E(I_{i-2,i-1}))' \quad = \quad \Phi(w_i) + (\mathbf{\Omega}^{*}_{<i}{}^{-1} \cdot \mathbf{\Omega}^{*}_{\boldsymbol{i},<\boldsymbol{i}})' \\ &(1-\Phi(w_1),\ ...1-\Phi(w_{i-1}), 1-\Phi_2(w_1, w_2, \rho_{12}), ..., 1-\Phi_2(w_{i-2}, w_{i-1}, \rho_{i-2,i-1}))' \end{aligned} \tag{5}$$

In above equation $\Phi_2(\cdot)$ is the bivariate CDF function. The other matrices, $\mathbf{\Omega}^{*}_{\boldsymbol{i},<\boldsymbol{i}}$ and $\mathbf{\Omega}^{*}_{<i}$, are constructed as follows:

$$\mathbf{\Omega}^{*}_{\boldsymbol{i},<\boldsymbol{i}} = \left(\mathbf{\Omega}_{\boldsymbol{i},<\boldsymbol{i}}, \mathbf{A}\right), \qquad \mathbf{\Omega}^{*}_{<i} = \begin{bmatrix} \mathbf{\Omega}^{*}_{<i} & \mathbf{B} \\ \mathbf{B}' & \mathbf{C} \end{bmatrix} \tag{6}$$

The composition of the three matrices **A**, **B** and **C** are given next. **A** is row vector of length (*i*–1)(*i*–2)/2, constituted by elements given by $\mathrm{Cov}(\tilde{I}_i, \tilde{I}_{jk})$ with $j < k < i$. **B** is matrix of size (*i*–1) × (*i*–1)(*i*–2)/2 and is constituted of elements given by $\mathrm{Cov}(\tilde{I}_l, \tilde{I}_{jk})$ with $1 < l < i$ *and* $j < k < i$. Finally, **C** is matrix of size (*i*–1)(*i*–2)/2 × (*i*–1)(*i*–2)/2 and is constituted of elements given by $\mathrm{Cov}(\tilde{I}_{jk}, \tilde{I}_{j'k'})$ with $j < k < i$ and $j' < k' < i$. Similar to the first order approximation we evaluate the two types of covariances involved in the creation of the matrices **A**, **B** and **C** as $\mathrm{Cov}(\tilde{I}_l, \tilde{I}_{jk}) = E\left(\tilde{I}_l \tilde{I}_{jk}\right) - E\left(\tilde{I}_l\right)E\left(\tilde{I}_{jk}\right)$ and $\mathrm{Cov}(\tilde{I}_{jk}, \tilde{I}_{j'k'}) = E\left(\tilde{I}_{jk}\tilde{I}_{j'k'}\right) - E\left(\tilde{I}_{jk}\right)E\left(\tilde{I}_{j'k'}\right)$. It can easily be observed that $E\left(\tilde{I}_l \tilde{I}_{jk}\right)$ will involve a trivariate CDF if all the three indices *l*, *j* and *k* are distinct. Similarly, $E\left(\tilde{I}_{jk}\tilde{I}_{j'k'}\right)$ evaluation will involve a four-variate CDF if all the four indices *i*, *j*, *i′* and *k′* are distinct. For the cases where the indices are not distinct the expectation can be evaluated using a

CDF function of a lower dimension depending on how many unique indices are present in the expectation function. The three matrices are constructed as given below:

$$\mathbf{A} = \left[\mathrm{Cov}(\tilde{I}_i,\tilde{I}_{12}) \quad \cdots \quad \mathrm{Cov}(\tilde{I}_i,\tilde{I}_{1,i-1}) \quad \mathrm{Cov}(\tilde{I}_i,\tilde{I}_{23}) \quad \cdots \quad \mathrm{Cov}(\tilde{I}_i,\tilde{I}_{i-2,i-1})\right] \tag{7}$$

$$\mathbf{B} = \begin{bmatrix} \mathrm{Cov}(\tilde{I}_1,\tilde{I}_{12}) & \cdots & \mathrm{Cov}(\tilde{I}_1,\tilde{I}_{1,i-1}) & \mathrm{Cov}(\tilde{I}_1,\tilde{I}_{23}) & \cdots & \mathrm{Cov}(\tilde{I}_1,\tilde{I}_{i-2,i-1}) \\ \mathrm{Cov}(\tilde{I}_2,\tilde{I}_{12}) & \cdots & \mathrm{Cov}(\tilde{I}_2,\tilde{I}_{1,i-1}) & \mathrm{Cov}(\tilde{I}_2,\tilde{I}_{23}) & \cdots & \mathrm{Cov}(\tilde{I}_2,\tilde{I}_{i-2,i-1}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\tilde{I}_{i-1},\tilde{I}_{12}) & \cdots & \mathrm{Cov}(\tilde{I}_{i-1},\tilde{I}_{1,i-1}) & \mathrm{Cov}(\tilde{I}_{i-1},\tilde{I}_{23}) & \cdots & \mathrm{Cov}(\tilde{I}_{i-1},\tilde{I}_{i-2,i-1}) \end{bmatrix} \tag{8}$$

$$\mathbf{C} = \begin{bmatrix} \mathrm{Cov}(\tilde{I}_{12},\tilde{I}_{12}) & \cdots & \mathrm{Cov}(\tilde{I}_{12},\tilde{I}_{1,i-1}) & \mathrm{Cov}(\tilde{I}_{12},\tilde{I}_{23}) & \cdots & \mathrm{Cov}(\tilde{I}_{12},\tilde{I}_{i-2,i-1}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\tilde{I}_{1,i-1},\tilde{I}_{12}) & \cdots & \mathrm{Cov}(\tilde{I}_{1,i-1},\tilde{I}_{1,i-1}) & \mathrm{Cov}(\tilde{I}_{1,i-1},\tilde{I}_{23}) & \cdots & \mathrm{Cov}(\tilde{I}_{1,i-1},\tilde{I}_{i-2,i-1}) \\ \mathrm{Cov}(\tilde{I}_{23},\tilde{I}_{12}) & \cdots & \mathrm{Cov}(\tilde{I}_{23},\tilde{I}_{1,i-1}) & \mathrm{Cov}(\tilde{I}_{23},\tilde{I}_{23}) & \cdots & \mathrm{Cov}(\tilde{I}_{23},\tilde{I}_{i-2,i-1}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\tilde{I}_{i-2,i-1},\tilde{I}_{12}) & \cdots & \mathrm{Cov}(\tilde{I}_{i-2,i-1},\tilde{I}_{1,i-1}) & \mathrm{Cov}(\tilde{I}_{i-2,i-1},\tilde{I}_{23}) & \cdots & \mathrm{Cov}(\tilde{I}_{i-2,i-1},\tilde{I}_{i-2,i-1}) \end{bmatrix} \tag{9}$$

An additional point to be noted in the second order approximation is that the initial decomposition of equation (1) can be done in three different ways, not taking into account the permutations of the approximation within a given decomposition. The three decompositions are using a bivariate, a trivariate or a four-variate CDF function for the first term of the decomposition and using the conditional probability approximation for the remaining dimensions. This can be written generically as given below:

$$\Pr(\boldsymbol{W} < \boldsymbol{w}) = \Pr(W_1 < w_1, \cdots W_n < w_n) \times \prod_{\mathrm{i=n+1}}^{\mathrm{I}} \Pr(W_i < w_i | W_1 < w_1,\ W_2 < w_2, ...,\ W_{i-1} < w_{i-1}) \tag{10}$$

The number of permutations that are possible depends on the value of *n* and is given by I!/n!. Joe (1995) uses *n*=4 for all the probability simulation experiments performed in that study; however, no reason is provided for selecting this value. We performed some simulation experiments for different values of *n* and observed that the accuracy levels of the probability, averaged across *I*!/*n*! permutation, were similar for all three values of *n*. The approximation where *n* takes the value 4 however takes lesser time, since there are fewer permutations, and this is the approximation scheme used in the later probability experiments. One exception to this is our first set of estimation experiments (discussed in

next section) because those set of experiments has only four dimensions of integration and the approximation where *n* equals four can be used only for cases where the dimension of integration is more than 4. Therefore for that particular experiment we use a value of three for *n*.

Another issue to be considered in the approximation is the method of evaluation of the trivariate and four-variate CDFs in the approximation formula. The first option is to use the standard multivariate cumulative normal probability algorithms such as the ones by Genz (1992), for computation of these CDFs; however, most of these are very time consuming. The other option is to use the first order MVNCD approximation of Joe (1995) for both the trivariate and four-variate CDFs. Using a mixture of both of these is the third option. Many preliminary experiments were performed and it was found that it was necessary to use a very accurate probability evaluation for the trivariate CDF and the first order MVNCD approximation of Joe (1995) was not sufficient. Finally, the trivariate CDF approximation given by Genz (2004) and the second order MVNCD approximation of Joe (1995) (presented in this study) for the four-variate CDF was found to perform the best. Note that the second order approximation of a four-variate CDF requires only a trivariate CDF and lower order CDFs.

6.2 EXPERIMENTAL DESIGN

We perform simulation experiments for four different types of model structures to evaluate and compare the performance of the second order MACML method (the one that uses the second order MVNCD) vis-à-vis the first order MACML method (the one that uses the first order MVNCD). The first two model structures will be based on a normally distributed random coefficient multinomial probit model with the first being a cross-sectional model and the second being a panel model. The subsequent two model structures will be based on a cross-sectional skew-normally distributed random coefficient multinomial probit models. For each of the model structures, 20 datasets are generated based on the underlying model structure and the assumed ‘true’ parameter values. The estimations are performed by the first order MACML method called henceforth as the MACML-FO method and the second order MACML method called

henceforth as the MACML-SO method. The estimations are performed ten times (for different permutations in the MACML-MVNCD decompositions) for each of the 20 datasets to get 10 different estimates each. The details of the simulation setup such as the number of observations and the parameter value used in the data generation vary in the four set of experiments and will be provided in detail in each of the relevant sub-sections.

6.2.1 Cross-Sectional Normal Random Coefficients Model Structure

The cross-sectional normally distributed random coefficient model structure with diagonal covariance specification considered in the simulation experiments of chapter 2 in section 2.2.1 is used here. For this model structure, the coefficient vector $\boldsymbol{\beta}_q$ is assumed to have a mean vector given by $\boldsymbol{b}$ = (1.5, –1, 2, 1, –2) and a covariance matrix given by $\boldsymbol{\Omega}$, a diagonal covariance matrix comprised of one along the diagonals and zero along the off-diagonals (same as in chapter 2). A random realization based on this structure is used to draw coefficients for each of the individuals. A dataset of 5000 observations is generated using this individual specific coefficient vector, exogenous variable values drawn from a standard normal distribution and an error vector also drawn from a standard normal distribution. Note that the error can have a more general covariance matrix structure subject to identification constraints; however, we use an i.i.d error structure in these experiments.

6.2.2 Panel Normal Random Coefficients Model Structure

We use the model setup that was used in section 2.2.2 of the second chapter. For this model also we generate the data using the same coefficient vector $\boldsymbol{\beta}_q$ that was assumed in the cross-sectional experiment, which was the mean vector given by $\boldsymbol{b}$ = (1.5, –1, 2, 1, –2) and a covariance matrix given by $\boldsymbol{\Omega}$, a diagonal covariance matrix comprised of one along the diagonals and zero along the off-diagonals. 800 individuals characteristics are created by drawing from the distribution of $\boldsymbol{\beta}_q$ and for each of these individuals 3 choice occasions are created for a total of 2400 observations.

6.2.3 Skew-Normal Random coefficient Model Structure

The details of formulation of this model can be obtained from section 5.3.1 of the fifth chapter. For this model structure we perform two sets of simulation experiments; the first set of experiments where the skewness is assumed to be small and the second set of experiments where the skewness is assumed to be higher. Following are the parameter values assumed for the distribution of MVSN distributed coefficients:

Low Skewness Model

$$\mathbf{b} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1.25 \end{pmatrix}, \text{ and } \boldsymbol{\Omega}_{+}^{*} = \begin{pmatrix} 1 & -0.4 & -0.4 & -0.4 \\ -0.4 & 1 & 0.16 & 0.16 \\ -0.4 & 0.16 & 1 & 0.16 \\ -0.4 & 0.16 & 0.16 & 1 \end{pmatrix}. \tag{11}$$

High Skewness Model

$$\mathbf{b} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}, \quad \boldsymbol{\omega} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1.25 \end{pmatrix}, \text{ and } \boldsymbol{\Omega}_{+}^{*} = \begin{pmatrix} 1 & -0.7 & -0.7 & -0.7 \\ -0.7 & 1 & 0.49 & 0.49 \\ -0.7 & 0.49 & 1 & 0.49 \\ -0.7 & 0.49 & 0.49 & 1 \end{pmatrix}. \tag{12}$$

5000 observations are generated based on these coefficients for the two models: low skewness model and high skewness model.

6.3 PERFORMANCE MEASURES

The performance measures used to compare the two methods, MACML-FO and MACML-SO are the same as the ones used in the earlier simulation experiments. They are the absolute percentage bias (APB), the approximation standard error, the asymptotic standard error, the approximation adjusted asymptotic standard error, and the computational time. Section 5.1 in chapter 5 explains in detail the steps involved in the calculation of these performance measures.

6.4 RESULTS

The results of the simulation experiments are presented for each of the four model structures used in the study in the following sub-sections.

6.4.1 Cross-Sectional Normal Random Coefficients Model Structure

The results of the simulation experiments for this model are presented in Table 6.1. The MACML method retrieves parameters quite accurately as has been shown in chapter 2 and we observe the trend here in these results also. The APB across all parameters is only 2.5% for MACML-FO and 1.9% for MACML-SO. The APB values are better for the MACML-SO estimates than the MACML-FO estimates for 9 out of the 10 parameters. Additionally, the APB values of the mean (location parameter) of the $\boldsymbol{\beta}_q$ vector is smaller than the APB of the standard deviation of the $\boldsymbol{\beta}_q$ vector, indicating that the mean of the $\boldsymbol{\beta}_q$ vector is more influential in the objective function than the standard deviation of the $\boldsymbol{\beta}_q$ vector. The efficiency of estimation is good for both the MACML-FO estimator and the MACML-SO estimator, with the asymptotic error being only about one-eighth of the parameters values. Also, the average asymptotic standard error values from the MACML-FO and MACML-SO estimators are very close to each other(0.151 and 0.152 respectively), suggesting that their efficiencies are very similar. The approximation standard error, which shows the noise generated due to not using all the permutations in the MVNCD decomposition and instead using only one permutation of the MVNCD decomposition, are very different for the two methods. The MACML-SO outperforms MACML-FO significantly in this regard. The average approximation standard error is only 0.005 (3% of the average asymptotic standard error) for the MACML-SO estimator compared to 0.019 (12% of the asymptotic standard error) for the MACML-FO estimator. The approximation adjusted asymptotic standard error of the MACML-SO is slightly better at 0.152 compared to the MACML-FO estimator which has an average approximation adjusted asymptotic error of 0.153.

The average computation time taken (including the computation of the covariance matrix) for the MACML-FO method is 1.96 minutes compared to 24.3 minutes for the MACML-SO method. This means that the MACML-SO takes about 12 folds more time than the MACML-FO method for this particular model structure and dimension. The performance of MACML-FO can theoretically be improved by using multiple permutations to decompose the MVNCD function in the MACML method, which however comes with an increase in computation time. As we saw in the introduction, for an approximation of an I dimensional integral using the first order MVNCD approximations, $I!/2$ permutations are possible for the decomposition into conditional probabilities, yielding a slightly different approximation each time. Averaging the probability across these permutations can give a more accurate probability evaluation. We performed estimations in which all the permutations were considered for each observation in order to evaluate the improvement in the performance of the MACML-FO method.

The result of this experiment was as follows; the average APB across parameters is 1.9%, which when compared to the MACML-FO with 1 permutation APB of 2.5% is an improvement. Interestingly, this APB of MACML-FO with all permutations is the same as the one for MACML-SO with 1 permutation. The asymptotic standard error is however slightly higher for the MACML-FO method with all permutations at 0.154 (compared to 0.153 for MACML-FO with1 permutation and 0.152 for MACML-SO with 1 permutation). The average computational time of the MACML-FO method increases to 17.5 minutes from 1.96 minutes (1 permutation of MACML-FO) and is indeed comparable to the 24.4 minutes observed for the MACML-SO method. Overall, from the results of this set of simulation experiment results, it can be said that the MACML-SO method improves the estimation marginally, especially in reducing the approximation standard error; however, there are no significant differences in the asymptotic standard errors. The APB of MACML-SO method with 1 permutation for each probability evaluation is as good as the APB of MACML-FO method that uses all the permutations.

Table 6.1: Simulation results for the normally distributed Cross-sectional random coefficient model

Parameter	True Value	MACML Method (First Order)					MACML Method (Second Order)				
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute Percentage Bias	Asym. SE	Approx. SE	Approx. Adjusted Asym. SE	Mean Estimate	Absolute Percentage Bias	Asym. SE	Approx. SE	Approx. Adjusted Asym. SE
Mean values of the β vector											
$b1$	1.500	1.472	1.89%	0.167	0.022	0.169	1.483	1.11%	0.168	0.005	0.168
$b2$	-1.000	-0.976	2.42%	0.113	0.014	0.114	-0.984	1.62%	0.114	0.003	0.114
$b3$	2.000	1.940	2.98%	0.218	0.028	0.219	1.955	2.27%	0.218	0.007	0.218
$b4$	1.000	0.977	2.30%	0.114	0.014	0.114	0.985	1.54%	0.114	0.003	0.114
$b5$	-2.000	-1.960	1.98%	0.220	0.028	0.222	-1.976	1.22%	0.221	0.007	0.221
Standard deviations of the β vector											
$\sigma1$	1.000	0.958	4.20%	0.135	0.017	0.137	0.966	3.43%	0.136	0.004	0.136
$\sigma2$	1.000	0.984	1.64%	0.136	0.016	0.137	0.991	0.93%	0.137	0.004	0.137
$\sigma3$	1.000	0.941	5.95%	0.135	0.017	0.136	0.946	5.44%	0.136	0.004	0.136
$\sigma4$	1.000	0.982	1.75%	0.136	0.017	0.137	0.989	1.14%	0.136	0.004	0.136
$\sigma5$	1.000	1.002	0.16%	0.140	0.016	0.141	1.008	0.79%	0.141	0.004	0.141
Overall Mean Value Across Parameters		-	2.53%	0.151	0.019	0.153	-	1.95%	0.152	0.005	0.152
Mean Time		1.96					24.3				
Average log(CML)		-0.79887					-0.79852				

Table 6.2: Simulation results for the normally distributed Panel random coefficient model

Parameter	True Value	MACML Method (First Order)					MACML Method (Second Order)				
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Standard Error Estimates		
		Mean Estimate	Absolute Percentage Bias	Asym. SE	Approx. SE	Approx. Adjusted Asym. SE	Mean Estimate	Absolute Percentage Bias	Asym. SE	Approx. SE	Approx. Adjusted Asym. SE
Mean values of the β vector											
$b1$	1.500	1.494	0.40%	0.177	0.053	0.184	1.529	1.91%	0.142	0.051	0.151
$b2$	-1.000	-1.005	0.46%	0.125	0.036	0.130	-1.028	2.84%	0.102	0.034	0.108
$b3$	2.000	1.976	1.22%	0.230	0.070	0.240	2.021	1.07%	0.180	0.067	0.192
$b4$	1.000	0.988	1.21%	0.124	0.035	0.129	1.011	1.10%	0.103	0.033	0.108
$b5$	-2.000	-1.991	0.47%	0.230	0.071	0.241	-2.036	1.82%	0.179	0.068	0.191
Standard deviations of the β vector											
$\sigma1$	1.000	0.972	2.83%	0.140	0.040	0.146	1.000	0.04%	0.115	0.041	0.122
$\sigma2$	1.000	0.964	3.61%	0.139	0.040	0.144	0.995	0.46%	0.110	0.039	0.117
$\sigma3$	1.000	0.988	1.24%	0.146	0.041	0.151	1.019	1.87%	0.122	0.039	0.128
$\sigma4$	1.000	0.997	0.27%	0.141	0.041	0.147	1.029	2.91%	0.115	0.038	0.121
$\sigma5$	1.000	0.971	2.95%	0.143	0.038	0.149	0.998	0.25%	0.117	0.038	0.124
Overall Mean Value Across Parameters		-	1.47%	0.159	0.046	0.166	-	1.43%	0.128	0.045	0.136
Mean Time		5.75					455.1				
Average log(CML)		-1.529					-1.523				

6.4.2 Panel Normal Random Coefficients Model Structure

Table 6.2 shows the results of the simulation experiments performed for this model structure. The average APB across parameters is 1.47% for the MACML-FO method compared to 1.43% for the MACML-SO method. The efficiency of estimation is good for both the estimation methods, with the asymptotic standard error being less than one-eighth of the average parameter values, indicating good recovery. The MACML-SO method does however have a better efficiency with an average asymptotic standard error across parameters of 0.128 compared to 0.159 for the MACML-FO method. The approximation standard error for the MACML-SO method (0.045) does not show much improvement compared to the approximation standard error for the MACML-FO method (0.046), unlike the cross-sectional model considered in the previous section. This could be because, as the dimensionality of the integration increases, the noise generated from using only one permutation in the second order MVNCD decomposition also increases. However, in spite of this, the approximation adjusted standard error of the MACML-SO method (0.136) is still better than the approximation adjusted standard error for the MACML-FO method (0.166).

The average computational time taken for each estimation run is 455 minutes for the MACML-SO method compared to just under 6 minutes for the MACML-FO method. This shows that the increased accuracy in terms of lower APB and lower asymptotic standard error for the MACML-SO method comes at a computational cost. Looking at the results it is difficult to judge which method provides a better result, therefore, in order to make a fair comparison, we increased the number of permutations of MVNCD decomposition used in the MACML-FO approximation to 100 (note that at a dimension of 8, we can have 20,160 permutations and using all of them will be computationally impractical). The average APB surprisingly increased from 1.5% to 3.4% for the MACML-FO method in spite of an increase in the number of permutations used. The asymptotic standard error also increased to 0.181 from 0.166 indicating a decrease in the efficiency, while the computational time increased to 509 minutes from 6 minutes. This shows that using more permutations and getting an average value for the probability for

the MACML-FO method does not ensure an improvement in the performance measures. Comparing these results with the MACML-SO results, it can be concluded that the MACML-SO method with 1 permutation outperforms the MACML-FO method with 100 permutations. This becomes clear looking at the following comparisons: APB 1.4% (MACML-SO) compared to 3.4% (MACML-FO with 100 permutations), asymptotic standard error of 0.136 (MACML-SO) compared to 0.181 (MACML-FO with 100 permutations) and finally the computational time of 455 minutes (MACML-SO) compared to 509 minutes (MACML-FO with 100 permutations).

6.4.3 Skew-Normal Random coefficient Model Structure – Low skewness

Table 6.3 shows the results of the simulation runs for the skew-normal random coefficient model with a low level of skewness in the skew distribution (ρ of -0.4). There are significant differences in the performance of the MACML-FO and MACML-SO methods. This could be due to both a higher dimensionality of 5 compared to the dimension of the cross-sectional model of 4 and also due to the higher correlation values in the correlation matrices of the MVNCD function, that are observed in the skew-normally distributed models. The MVNCD approximations are known to deteriorate with an increase in correlation values (Joe, 1995) and the two methods might have different accuracies to higher correlation values and hence the difference in the observed performance measures. The average APB across parameters for the MACML-FO method at 15.1% is significantly higher compared to the MACML-SO method which has an average APB of only 6.4%. Interestingly, most of this bias is coming from the parameters related to the coefficients that are skew-normally distributed. The average APB of the parameters related to the skew-normally distributed coefficients is 21.3% for the MACML-FO estimation. The corresponding APB for these parameters from the MAMCL-SO estimation is only 8%. There is a consistent upward bias of around 35% to 40% in the skewness estimates of MACML-FO, unlike the MACML-SO estimates, where there is no such trend. On the other hand, there are no significant differences in the APB of the parameters related to the normally distributed coefficients between the two estimation methods (3.9% for MACML-FO and 3.7% for MACML-SO). The efficiency

is reasonably good with the standard error being on an average only one-sixth of the parameter values for the MACML-FO estimates and one-fifth for the MACML-SO estimates. The average asymptotic error is 0.117 for the MAMCL-FO method which is considerably smaller than the average asymptotic error of 0.209 for the MACML-SO method. The average approximation error is also smaller for the MAMCL-FO (0.090) than the MAMCL-SO method (0.105). From these two statistic it is clear that the approximation adjusted standard error of the MACML-FO should be better than the approximation adjusted standard error for the MAMCL-SO method, which is indeed the case (0.151 for MACML-FO compared to 0.238 for MACML-SO). From this comparison it might seem that MACML-FO is more efficient than MAMCL-SO; however, the consistent biases in the parameter estimates of MACML-FO negate these apparent efficiency gains.

Similar to earlier two models, the MACML-SO takes more computational time (120 minutes) than the MACML-FO method (5.5 minutes). To make a fairer computational time comparison, the number of permutations in the MVNCD decomposition was increased to the maximum possible for the 5 dimensional CDF, which is 60. This increased the average computational time to 271 minutes. For this model also, there was deterioration in the average APB as it increased from 15% to 24%. The APBs of the skewness parameters were particularly high with upward biases of up to 65%. The asymptotic standard error was, however, lower at 0.108 compared to 0.151. Overall, it can be said that the MACML-SO method performs better than the MACML-FO method for the skew-normal model (with low skewness), especially in decreasing the APB of the estimated parameters. In addition to that, increasing the number of permutations in the MACML-FO seems to impact the estimation results adversely.

Table 6.3: Simulation results for the skew-normally distributed random coefficient model with low skewness

		MACML Method (FO)					MACML Method (SO)				
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Standard Error Estimates		
Parameter	True Value	Mean Estimate	Absolute Percentage Bias	Asym. SE	Approx. SE	Approx. Adj. Asym. SE	Mean Estimate	Absolute Percentage Bias	Asym. SE	Approx. SE	Approx. Adj. Asym. SE
Location parameters of the β vector											
b1	-1.000	-0.874	12.6%	0.126	0.123	0.176	-1.038	3.8%	0.294	0.170	0.339
b2	-1.000	-0.873	12.7%	0.124	0.133	0.182	-0.992	0.8%	0.292	0.185	0.346
b3	-1.000	-0.878	12.2%	0.136	0.111	0.176	-0.935	6.5%	0.358	0.252	0.438
Skewness parameters of the β vector											
ρ1	-0.400	-0.567	41.9%	0.069	0.104	0.125	-0.317	20.6%	0.258	0.159	0.303
ρ2	-0.400	-0.557	39.2%	0.066	0.118	0.135	-0.403	0.7%	0.265	0.175	0.317
ρ3	-0.400	-0.544	35.9%	0.067	0.084	0.108	-0.449	12.2%	0.245	0.176	0.302
Scale parameters of the β vector											
ω1	1.000	1.132	13.2%	0.150	0.097	0.179	1.092	9.2%	0.185	0.080	0.202
ω2	1.000	1.142	14.2%	0.152	0.098	0.180	1.088	8.8%	0.187	0.092	0.208
ω3	1.250	1.376	10.1%	0.172	0.112	0.205	1.364	9.1%	0.239	0.123	0.268
Mean values of the γ vector											
c1	1.000	1.023	2.3%	0.114	0.058	0.128	1.024	2.4%	0.120	0.013	0.120
c2	1.000	1.023	2.3%	0.114	0.057	0.127	1.024	2.4%	0.119	0.013	0.120
Covariance elements of the γ vector											
σ1	1.000	1.050	5.0%	0.137	0.065	0.152	1.052	5.2%	0.143	0.015	0.144
Σ12	0.500	0.533	6.5%	0.075	0.027	0.080	0.521	4.3%	0.076	0.004	0.076
σ2	1.000	1.036	3.6%	0.135	0.072	0.154	1.041	4.1%	0.142	0.015	0.143
Overall Mean Value Across Parameters			15.1%	0.117	0.090	0.151		6.4%	0.209	0.105	0.238
Mean Time		5.5					120.0				

Table 6.4: Simulation results for the skew-normally distributed random coefficient model with high skewness

		MACML Method (FO)					MACML Method (SO)				
		Parameter Estimates		Standard Error Estimates			Parameter Estimates		Standard Error Estimates		
Parameter	True Value	Mean Estimate	Absolute Percentage Bias	Asym. SE	Approx. SE	Approx. Adj. Asym. SE	Mean Estimate	Absolute Percentage Bias	Asym. SE	Approx. SE	Approx. Adj. Asym. SE
Location parameters of the β vector											
b1	-1.000	-0.948	5.2%	0.117	0.084	0.144	-1.095	9.5%	0.260	0.117	0.285
b2	-1.000	-0.937	6.3%	0.117	0.090	0.148	-1.120	12.0%	0.247	0.116	0.273
b3	-1.000	-0.984	1.6%	0.125	0.099	0.159	-1.116	11.6%	0.309	0.133	0.337
Skewness parameters of the β vector											
ρ1	-0.700	-0.798	14.0%	0.043	0.050	0.066	-0.615	12.2%	0.203	0.096	0.225
ρ2	-0.700	-0.797	13.8%	0.042	0.049	0.064	-0.594	15.1%	0.210	0.098	0.232
ρ3	-0.700	-0.767	9.5%	0.040	0.046	0.062	-0.617	11.8%	0.201	0.089	0.219
Scale parameters of the β vector											
ω1	1.000	1.149	14.9%	0.146	0.117	0.187	1.007	0.7%	0.210	0.081	0.225
ω2	1.000	1.166	16.6%	0.146	0.106	0.180	0.999	0.1%	0.191	0.078	0.206
ω3	1.250	1.382	10.5%	0.166	0.107	0.198	1.260	0.8%	0.247	0.093	0.264
Mean values of the γ vector											
c1	1.000	1.044	4.4%	0.109	0.063	0.126	1.034	3.4%	0.120	0.011	0.120
c2	1.000	1.046	4.6%	0.110	0.063	0.127	1.037	3.7%	0.120	0.011	0.121
Covariance elements of the γ vector											
σ1	1.000	1.064	6.4%	0.129	0.075	0.149	1.046	4.6%	0.141	0.013	0.142
Σ12	0.500	0.508	1.6%	0.067	0.033	0.075	0.478	4.4%	0.072	0.005	0.072
σ2	1.000	1.075	7.5%	0.130	0.074	0.149	1.062	6.2%	0.142	0.012	0.143
Overall Mean Value Across Parameters			8.4%	0.106	0.076	0.131		6.9%	0.191	0.068	0.205
Mean Time		5.4					109.5				

6.4.4 Skew-Normal Random coefficient Model Structure – High skewness

The results of the simulation experiments for the skew-normal random coefficient model with high skewness (ρ of -0.7) are presented in Table 6.4. The average APB values across parameters is only marginally better for the MACML-SO method (6.9%) compared to the MACML-FO method (8.4%), unlike the results that we saw in the previous skew-normal model with low skewness value. However, there are significant differences in the APB values for various sets of parameters. While the location parameters of the β vector have lower average APB for the MACML-FO (4.4%), the same for the MACML-SO method is considerably higher at 11.0%. On the other hand, there is high bias in the scale parameters of the MAMCL-FO method with an average APB of 14%, while the MAMCL-SO method has almost no bias (0.5%). The last set of parameters related to the β vector, the skewness parameters, have similar APB values for the two estimation method; however, the biases are in different directions. The parameters estimated using the MAMCL-FO method has an upward bias whereas the parameters estimated using the MAMCL-SO method has a downward bias. Finally, the APB of the parameters related to the γ vector, show no significant differences in the APB values for the MACML-FO and MACML-SO method, which is similar to what was observed for the simulation experiments for the previous skew-normal random coefficient model with low skewness. The efficiency of estimation is reasonably good as seen from the average standard errors that are less than a fifth of the parameter values. Similar to the previous simulation experiment, the MACML-FO estimates have a lower average asymptotic error (0.106) than the asymptotic error observed for the MACML-SO estimates (0.191). The approximation standard errors are similar for both the estimation methods and the approximation adjusted asymptotic standard error are as expected better for the MACML-FO method. The MACML-FO has an average approximation adjusted asymptotic standard error of 0.131 compared to the average approximation adjusted asymptotic standard error of 0.205 for the MAMCL-SO method.

Computation run times are similar to the previous experiment with the MACML-FO method taking only an average of 5.4 minutes compared to the MACML-SO

method's average runtime of 110 minutes. Similar to earlier experiments, the number of permutations used for each probability evaluation was increased to the maximum for 5 dimension, which is 60, and the models where estimated using the MACML-FO method. The APB increased from 8.4% to 10.4% while the average asymptotic standard error decreased from 0.131 to 0.105. The computational time increased from 5.4 minutes to 268 minutes. Overall, it maybe concluded that for this model also the MACML-SO method outperforms the MACML-FO method. The APB values from the MACML-SO method are slightly better than the same from the MACML-FO method and moreover, the MACML-FO method performance seems to deteriorate with an increase in the number of permutations used in each probability evaluation.

6.5 CONCLUSIONS

The simulation experiments of the second chapter performed on normally distributed random coefficient models showed that the MAMCL method performs extremely well in recovering parameters and outperforms the MSL method in all the performance evaluation measures considered. Since the introduction of the MACML estimation method, the MACML method has been used for a variety of applications to estimate different types of models. MACML method has been used in applications such as land use change, household work arrangement choices, school mode choice etc. The models structures estimated range from simple multinomial probit models to unordered models with spatial lag structures. Further as showed in the fifth chapter of this dissertation the MACML estimation approach can be used for the estimation a non-normal mixing distribution, namely a generalized normal distribution called as the skew-normal distribution. The simulation experiments performed for this model structure indicated that the parameters are recovered reasonably well. While the performance measures observed for the skew-normal random coefficient model's simulation experiments were reasonably good, they were not as good as the ones observed for the normally distributed random coefficient models.

The MACML approach of Bhat (2011a) uses a first order MVNCD function approximation proposed in Joe (1995). This first order approximation requires only univariate and bivariate normal CDF functions to evaluate and is therefore extremely fast and easy to implement. The first order approximation is however not very accurate at an individual probability evaluation, as simulation experiments for probability computations in Joe (1995) demonstrates. However, since the MACML estimation involves multiple probability evaluation of the order of hundreds or thousands of observations, the individual inaccuracies are not very critical, as shown by very good parameter recovery for normally distributed random coefficient models and reasonably good parameter recovery of skew-normally distributed random coefficient models. The MVNCD approximations are known to deteriorate as the elements in correlation matrix takes values closer to 1 or closer to -1 (Joe, 1995). This could be one of the reasons for the poorer performance of the MACML method in estimating skew-normal models where due to the high correlation values that occur in the multivariate probability expressions of the skew-normal models. In this study we incorporate the second order MVNCD approximation proposed in Joe (1995) within the MACML estimation approach (call this the second order MACML approach or the MACML-SO) and evaluate its performance in parameter estimation for a variety of models. The model structures considered in simulation study include the normally distributed random coefficient models for both a cross-sectional and panel case, and a skew-normally distributed cross-sectional random coefficient model.

The results of the study indicate that the MACML-SO method outperforms the MACML-FO method. However, the degree to which the MACML-SO method performs better is different in the four models that were considered in this study. In the normally distributed cross-section random coefficient model, the performance of the MACML-SO method was only marginally better and the difference vanished once the number of permutation used in the probability evaluation was increased to the maximum possible for the MAMCL-FO method. On the other hand, the results of the normally distributed panel random coefficient model clearly show the better performance of the MACML-SO

method, especially in having a much better efficiency. In addition to that the performance of the MACML-FO method deteriorated as the number of permutations used in the probability evaluation was increased to hundred (from one), as shown by an increase in the APB from 1.5% to 3.4% and asymptotic standard error from 0.166 to 0.181. The results from the skew-normal distributed random coefficient models indicate that the MAMCL-SO method does a much better job at recovering parameters than the MAMCL-FO method. For the skew-normal model with low skewness (ρ of -0.4), the APB was 15.1% for the MACML-FO estimates compared to 6.4% for the MACML-SO estimates. Similarly, for the skew-normal model with high skewness (ρ of -0.7), the APB was 8.4% for the MACML-FO method compared to 6.9% for the MAMCL-SO method. However, the approximation adjusted standard errors were higher for the MACML-SO estimates compared to the MACML-FO estimates in both the cases. For the skew-normal models also the APBs of MACML-FO estimator increased when the number of permutations were increased from one permutation to all possible (60) permutations. The APB for the low skewness model increased from 15% to 24% while the APB of the high skewness model increased from 8.4% to 10.4%. Even though with only one permutation, the MACML-FO estimations are much faster, the computational time increases and becomes comparable or more than the computational time for the MACML-SO estimations when the permutations are increased. In conclusion it may be said that when more permutations are used for the MACML-FO estimation, the results improve only for the normally distributed cross-sectional random coefficient model. For the other three models the results deteriorate as the number of permutations are increased and is in general worse than the results from the MACML-SO estimator. At the same time the MACML-SO method takes less time than the MACML-FO method with all (or a large number) the permutations.

To conclude, this chapter proposes the improvement of the MACML estimation approach introduced in Bhat (2011a) by using a second order approximation for the MVNCD function used in the MACML approach. The second order MVNCD function uses the trivariate and four-variate normal CDFs in addition to the univariate and

bivariate CDFs and is therefore more time consuming. The results from the simulation experiments show that the MACML-SO method performs better than the MACML-FO method. At the same time, results also informs us that the MACML-FO method does not necessarily give better results when number of permutations used in the MVNCD function evaluation is increased. Therefore, the MACML-SO method is one way to go if improved accuracy is required in estimations.

CHAPTER 7: CONCLUSIONS

7.1 SUMMARY

Multinomial Logit model (MNL) is one of the most widely used models in the area of transportation for discrete choice modeling. The simple closed form expression for the choice probability in an MNL model makes the estimation easy to implement and computationally fast. However, the MNL model has several disadvantages such as the familiar independence from irrelevant alternatives property and the inflexible substitution patterns implied by the model structure. In order to overcome these shortcomings, researchers have used alternative models that are more flexible such as the mixed multinomial logit model (MMNL) and the multinomial probit model (MNP), which are in turn estimated using maximum simulated likelihood (MSL) methods. In spite of the advent of these advanced models, applications of these models in the literature are mostly to problems with low dimensions due to the high and often impractical estimation time encountered as the dimension of the problem increases. The Maximum Approximate Composite Marginal Likelihood (MACML) estimation approach has, however, provided an opportunity to the application of discrete choice models to more complex and higher dimension problems. In this dissertation we first evaluate and compare the ability of the MACML method and the MSL method to estimate/recover parameters using simulation experiments for a variety of different model structures. Second, the MACML method is employed to estimate a children's school mode choice with spatial interaction. Third, a land use change problem is formulated within a spatially explicit economic structural framework and is estimated using the MACML approach. Fourth, we propose the use of the multivariate skew-normal distribution function to accommodate non-normal mixing in MNP models and demonstrate the ability of MACML approach to estimate such models. Finally, an improvement to the MACML method is proposed by the incorporation of second order MVNCD approximation.

In the second chapter of this dissertation, we evaluated the ability of the MACML estimation approach to recover parameters from finite samples in mixed cross-sectional and panel multinomial probit models. Comparisons with the results from MSL estimation

approach indicate that the MACML approach recovers parameters much more accurately than the MSL approach in all model structures and covariance specifications. The MACML inference approach also estimates the parameters efficiently, with the asymptotic standard errors being, in general, only a small proportion of the true values. As importantly, the MACML inference approach takes only a very small fraction of the time needed for MSL estimation. As the number of alternatives in the unordered-response model increases, one can expect even higher computational efficiency factors for the MACML over the MSL approach. Further, as should be evident in the panel intra- and inter-individual random coefficients case, the MSL is all but practically infeasible when the mixing structure leads to an explosion in the dimensionality of integration in the likelihood function, but these situations are handled with ease in the MACML approach.

In the third chapter, a school mode choice model that is capable of capturing the unobserved spatial interaction effects that may potentially influence household decision-making processes when choosing a mode of transportation for children's trips to and from school is formulated and estimated. The estimation is done on a sample of children residing in Southern California whose households responded to the 2009 National Household Travel Survey in the United States. It is found that spatial correlation effects are statistically significant, and that these effects arise from interactions among households that are geographically close to one another.

In the fourth chapter, an empirical land-use model is formulated within a spatially explicit economic structural framework. The underlying framework goes beyond mechanistic fitting models for the spatial process of land use change to more closely link landowner decision behavior to land use patterns. At the same time, the formulation explicitly considers spatial "spillover" effects in the decisions of land-owners of proximately located parcels, heterogeneity in the decision-making process of different land owners and stationary across-time correlation in land uses for the same spatial unit. The analysis is undertaken using the City of Austin parcel-level land use database for multiple years (1995, 2000, 2003, and 2006). The estimation results indicate that proximity to highways and other roadways, distance from flood plains, parcel location in

the context of existing development, and distance from schools are all important determinants of land-use. As importantly, the results provide very strong evidence of temporal dependency and spatial dynamics in land-use decisions. There is also a suggestion that major highways may not only physically partition regions, but may also act as social barriers for dyadic interactions among individuals.

In the fifth chapter, we propose the use of the multivariate skew-normal distribution function to accommodate non-normal mixing in cross-sectional and panel multinomial probit (MNP) models. The combination of skew-normal mixing and the MNP kernel lends itself nicely to estimation using the MACML approach. Simulation results for the cross-sectional case show that our proposed approach does well in recovering the underlying parameters, and also highlights the pitfalls of ignoring non-normality of the continuous mixing distribution when such non-normality is present. At the same time, the proposed model obviates the need to assume a pre-specified parametric distribution for the mixing, and allows the estimation of a very flexible, but still parsimonious, mixing distribution form.

In the sixth chapter, we improve the accuracy of the MACML method by extending the MVNCD Function approximation used in the MACML approach to a second order approximation. The results from the simulation experiments indicate that the second order MACML approach improves the estimation by giving a lower average percentage bias (APB) across parameters in all the models. The second order MACML gives lower asymptotic standard errors for the normally distributed random coefficient models but gives higher asymptotic standard errors for the skew-normally distributed random coefficient models. However, the higher APB values in first order MACML results for the skew-normal random coefficient models negate the observed lower asymptotic errors. Both the methods of estimations are based on probability decomposition that can be done in multiple permutations and hence averaging across all the permutations is one way to improve accuracy. The number of permutations used in each probability computation in the first order MACML approach was increased from one to the maximum (or a very high number) in order to evaluate the performance

changes. It was observed that except for the cross-sectional normally distributed random coefficient model, the performance of the first order MACML method (surprisingly) deteriorated with an increase in the number of permutations. Thus the second order MACML method outperformed the first order MACML method with multiple permutations while at the same taking less computation time for estimation.

7.2 FUTURE RESEARCH EXTENSIONS

The different studies undertaken in this dissertation can be extended in several ways. These possible future research issues will be discussed below for each of the chapters. Some of these avenues of research try to relax the assumptions used in the models while others try to generalize or make estimation faster and efficient.

For the simulation experiments of the second chapter, it would be useful to undertake a similar analysis as the one there for varying numbers of random coefficients (say, 10 and 20 random coefficients) to examine the effectiveness of the MACML approach with different numbers of random coefficients. Conceptually, this should not have much of an effect on the MACML procedure, but the empirical evidence needs to be generated. Second, the MACML likelihood procedure and the gradient procedure have been coded in the GAUSS matrix programming language. Currently, a scalar version of the MVNCD approximation procedure has been coded, which implies that the MACML code calls the MVNCD code each time a MVNCD function is to be approximated. This MACML approach can be speeded up by vectorizing the MVNCD approximation procedure, so that the procedure returns the approximated values for multiple MVNCD evaluations at once. Third, the MVNCD procedure is written to cycle through until a permutation is used that provides a non-negative value for the MVNCD function evaluation. That is, it is possible that the first permutation leads to a value for the MVNCD function approximation that is negative, in which case the code automatically attempts a different permutation for the decomposition of the MVNCD function into marginal and conditional probabilities. This situation is relatively rare, but can happen as the gradient procedure searches for an update direction. In such situations, one may seek a different permutation (as we have done) or just increase the number of permutations to

make the approximation. These, and potentially other automated techniques, can be compared in future research.

In the third chapter, the study accommodates spatial dependence due to proximity in residential locations of children and social interaction effects. An avenue for future research would be to extend the dependence effects to include proximity in school locations of children, with the notion that peer effects at school may also impact children's school mode choice. This additional effect can be accommodated in a straightforward manner in the proposed methodology by defining another weight matrix W_k that corresponds to school location proximity, and considering this weight matrix as one additional finite mixture dimension affecting the overall weight matrix W ($W = \sum_{k=1}^{K} \varphi_k W_k$). However, this would require the identification of the schools that each child in the sample goes to, with a geo-coding of these school locations. This information is not available in the NHTS data used in the study, but may be available in other activity-travel data sets in which each activity episode location is geo-coded.

The fourth chapter that deals with the land use change analysis can also be extended in several ways, including a more rigorous theoretical and simulation-based evaluation of estimator efficiency related to the specification of the composite marginal likelihood function, consideration of more flexible forms of spatial modeling that combine spatial lag and spatial error formulations, and the incorporation of additional parcel-level, pedo-climatic, and regional-level externalities. Another area that can be explored with regard to the empirical analysis is the impact of scale of analysis on results, which goes back to the well known issue of the modifiable areal unit problem (MAUP). In the analysis performed in the study, we considered an area of size 242m×242m as a single unit. This could be varied to both smaller and larger units. On the other hand each individual parcels can be treated as an independent spatial unit. Comparison of results from such different analysis can throw more light on the robustness of results and effects of different factors at different scales.

The skew-normal distribution introduced in chapter 5 in the context of discrete choice models can be extended to more advanced discrete choice models. However, additional simulation experiments needs to be performed to examine the effectiveness of the approach in settings with spatial dependencies and social dependencies across decision units, and combinations of temporal, spatial, and social dependencies. Application of the skew-normal random coefficient models to empirical problems such as the mode choice problem and comparison of results with random coefficient models with lognormal distribution is another avenue of research.

Finally the second order extension of the MACML approach of the sixth chapter can be further extended. First, the performance of the two methods (first order and second order) can be evaluated for other model structures such as the spatial lag model. Second, the second order approximation used in this study uses the very accurate approximation for the trivariate CDF and uses the second order MVNCD approximation for the four-variate CDF. This decreases the computational time significantly. However, improved accuracy can be achieved by instead using an accurate approximation for the four-variate CDF function in the second order MVNCD approximation. This could be evaluated using a similar simulation experiment.

APPENDIX A

Using the notations in Section 5.2, the moments of the SSN distribution are most easily obtained from the moment generating function of Z, which is given by:

$$
\begin{aligned}
M(t) = E[\exp(tz)] &= \int_{z=-\infty}^{\infty} \exp(tz)\,\phi(z)\,\Phi(\alpha z)\,dz \\
&= 2\int_{z=-\infty}^{\infty} \exp(tz)\,\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-z^2}{2}\right)\Phi(\alpha z)\,dz \\
&= 2\int_{z=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left\{z^2-2tz+t^2\right\}+\frac{t^2}{2}\right)\Phi(\alpha z)\,dz \\
&= 2\int_{z=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(\frac{t^2}{2}\right)\exp\left(-\frac{1}{2}(z-t)^2\right)\Phi(\alpha z)\,dz \\
&= 2\exp\left(\frac{t^2}{2}\right)\int_{z=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(z-t)^2\right)\Phi(\alpha z)\,dz \\
&= 2\exp\left(\frac{t^2}{2}\right)\int_{u=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)\Phi(\alpha(u+t))\,du\,, \text{ where } u = z\text{-}t \\
&= 2\exp\left(\frac{t^2}{2}\right)\int_{u=-\infty}^{\infty} \phi(\mathrm{u})\,\Phi(\alpha(u+t))\,du \\
&= 2\exp\left(\frac{t^2}{2}\right)\mathrm{E}\left(\Phi(\alpha u+\alpha t)\right) \\
&= 2\exp\left(\frac{t^2}{2}\right)\Phi\left(\frac{\alpha t}{\sqrt{1+\alpha^2}}\right) \text{ using } \mathrm{E}\left\{\Phi(\alpha u+\alpha t)\right\} = \Phi\left(\frac{\alpha t}{\sqrt{1+\alpha^2}}\right) \\
&\qquad\qquad \text{(from Ellison, 1964)}
\end{aligned}
\tag{A.1}
$$

.

In the above expression, $\rho = \dfrac{\alpha}{\sqrt{1+\alpha^2}}$. From above, the first three moments of the distribution may be written as follows with $b = \sqrt{2/\pi}$:

$$E(Z) = \mu_Z = \left.\frac{dM(t)}{dt}\right|_{t=0} = b\rho; \quad \mathrm{Var}(Z) = \sigma_Z^2 = \left.\frac{d^2M(t)}{dt^2}\right|_{t=0} = 1-b^2\rho^2, \text{ and}$$

$$\mathrm{Skew}(Z) = \gamma_Z = \left.\frac{d^3M(t)}{dt^3}\right|_{t=0} = \left(\frac{4-\pi}{2}\right)^2\left(\frac{\mu_Z^2}{\sigma_Z^2}\right)^{3/2},$$

where γ_Z is the Pearson index of skewness that is a measure of asymmetry. When $\alpha = 0$, γ_Z=0 as should be the case for the normal distribution. The moments for the variable $Y = \xi + \omega Z$, which is non-standard skew-normally distributed, may be obtained as $\mu_Y = \xi + \omega\mu_Z$, $\sigma_Y^2 = \omega^2(1-\sigma_Z^2)$, and $\gamma_Y = \gamma_Z$.

From Equation (5) of chapter 5,

$$\begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Then,

$$f(M_1, M_2 \mid M_1 > 0) = \frac{\phi_2(M_1, M_2)}{\text{Prob}[M_1 > 0]} = 2\,\phi_2(M_1, M_2),$$

where ϕ_2 is the standard bivariate normal density function.

$$\begin{aligned} f(M_2 \mid M_1 > 0) &= \int_{M_1=0}^{\infty} 2\,\phi_2(M_1, M_2)\,dM_1 \\ &= 2\int_{M_1=0}^{\infty} \phi(M_2)\frac{1}{\sqrt{1-\rho^2}}\phi\left[\frac{M_1 - \rho M_2}{\sqrt{1-\rho^2}}\right] dM_1 \\ &= 2\,\phi(M_2)\left[1 - \Phi(-\alpha M_2)\right], \text{ where } \alpha = \frac{\rho}{\sqrt{1-\rho^2}}, \end{aligned} \tag{A.2}$$

or $f(z) = \tilde{\phi}(z;\alpha) = 2\,\phi(z)\,\Phi(\alpha z)$

The moment generating function of $\mathbf{Z}$ is: $M_Z(\mathbf{t}) = 2\exp\left(\frac{1}{2}\mathbf{t}'\boldsymbol{\Omega}^*\mathbf{t}\right)\Phi(\boldsymbol{\rho}'\mathbf{t})$.

The first three moments of the distribution may subsequently be obtained from the function above in a straightforward fashion with $b = \sqrt{2/\pi}$:

$E(\mathbf{Z}) = \boldsymbol{\mu}_Z = b\boldsymbol{\rho}$; $\text{Var}(\mathbf{Z}) = \boldsymbol{\Omega}^* - \boldsymbol{\mu}_\mathbf{Z}\boldsymbol{\mu}_\mathbf{Z}'$, and

$$\text{Skew}(\mathbf{Z}) = \boldsymbol{\gamma}_Z = \left(\frac{4-\pi}{2}\right)^2 \left(\frac{\boldsymbol{\mu}_\mathbf{Z}'(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{\mu}_\mathbf{Z}}{\mathbf{1} - \boldsymbol{\mu}_\mathbf{Z}'(\boldsymbol{\Omega}^*)^{-1}\boldsymbol{\mu}_\mathbf{Z}}\right)^3,$$

The moments for the variable $\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z}$, which is non-standard skew-normally distributed, may be obtained as $\boldsymbol{\mu}_\mathbf{Y} = \boldsymbol{\xi} + \boldsymbol{\omega}\boldsymbol{\mu}_\mathbf{Z}$, $\text{Var}(\mathbf{Y}) = \boldsymbol{\omega}\mathbf{Var(Z)}\boldsymbol{\omega}$, and $\boldsymbol{\gamma}_Y = \boldsymbol{\gamma}_Z$. For

future reference, we will also write the moment generating function of **Y** (obtained from Equation (11) of chapter 5) as follows:

$$M_Y(\mathbf{t}) = E[\exp(\mathbf{t}'\mathbf{Y})] = E(\exp[\mathbf{t}'(\boldsymbol{\xi} + \boldsymbol{\omega}\mathbf{Z})]) = \exp(\mathbf{t}'\boldsymbol{\xi})E(\mathbf{t}'\boldsymbol{\omega}\mathbf{Z})$$
$$= 2\exp\left(\boldsymbol{\xi}'\mathbf{t} + \frac{\mathbf{1}}{\mathbf{2}}\mathbf{t}'\boldsymbol{\Omega}\mathbf{t}\right)\Phi(\boldsymbol{\rho}'\boldsymbol{\omega}\,\mathbf{t}).$$

(A.3)

REFERENCES

Aigner, D.J., C.A.K. Lovell, and P. Schmidt (1977) Formulation and estimation of stochastic frontier production function model. *Journal of Econometrics,* 6(1), 21-37.

Albert, J.H., and S. Chib (1993) Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association,* 88(422), 669-679.

Amador, F.J., R. Gonzales, and J. Ortuzar (2005) Preference heterogeneity and willingness to pay for travel time savings. *Transportation* 32(6), 627-647.

Anas, A., R. Arnott, and K.A. Small (1998) Urban spatial structure. *Journal of Economic Literature* 36(3), 1426-1464.

Anselin, L. (1988) *Spatial econometrics: Methods and models*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Anselin, L. (2003) Spatial externalities, spatial multipliers and spatial econometrics. *International Regional Science Review*, 26(2), 153-166.

Anselin, L., (2006) Spatial econometrics. In: Mills T., Patterson K. (eds), *Palgrave Handbook of Econometrics: Volume 1, Econometric Theory,* Palgrave Macmillan Basingstoke, 901-969.

Arellano-Valle, R.B., and A. Azzalini (2006) On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* 33(3), 561-574.

Arellano-Valle, R.B., and A. Azzalini (2008) The centred parametrization for the multivariate skew-normal distribution. *Journal of Multivariate Analysis* 99(7), 1362-1382.

Arellano-Valle, R.B., and M.G. Genton (2005) On fundamental skew distributions. *Journal of Multivariate Analysis* 96(1), 93-116.

Azzalini, A., (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* 12(2), 171-178.

Azzalini, A., (2011) Selection models under generalized symmetry settings. *Annals of the Institute of Statistical Mathematics,* 64(4), 737-750

Azzalini, A., and A. Capitanio (1999) Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B,* 61(3) 579-602.

Azzalini, A., and A. Dalla Valle (1996) The multivariate skew-normal distribution. *Biometrika* 83(4), 715-726.

Balcombe, K., A. Chalak, and I.M. Fraser (2009) Model selection for the mixed logit with bayesian estimation. *Journal of Environmental Economics and Management*, 57(2), 226–237.

Bartels, R., D.G. Fiebig, and A. van Soest (2006) Consumers and experts: an econometric analysis of the demand for water heaters. *Empirical Economics* 31(2), 369-391.

Bastin, F., C. Cirillo, and P.L. Toint (2010) Estimating non-parametric random utility models, with an application to the value of time in heterogeneous populations. *Transportation Science*, 44(4) 537-549.

Beck, N., K.S. Gleditsch, and K. Beardsley (2006) Space is more than geography: using spatial econometrics in the study of political economy. *International Studies Quarterly*, 50(1), 27-44.

Beck, L.F. and A.I. Greenspan (2008) Why don't more children walk to school? *Journal of Safety Research*, 39, 449-452.

Beron, K.J., and W.P.M. Vijverberg (2004) Probit in a spatial context: a Monte Carlo analysis. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, Berlin, 169-196.

Bhat, C.R. (1995) A heteroscedastic extreme-value model of intercity mode choice. *Transportation Research Part B*, 29(6), 471-483.

Bhat, C.R. (1997a) Work travel mode choice and number of nonwork commute stops. *Transportation Research Part B*, 31(1), 41-54.

Bhat, C.R. (1997b) An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, 31(1), 34-48.

Bhat, C.R. (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B*, 37(9), 837-855.

Bhat, C.R. (2011a) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B*, 45(7), 923-939.

Bhat, C.R. (2011b). The MACML estimation of the normally-mixed multinomial logit model. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.

http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/MACML_Estim_Norm_MML_Model.pdf

Bhat, C.R., and S. Castelar (2002). A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B,* 36(7), 593-616.

Bhat, C.R. and J.Y. Guo (2004) A mixed spatially correlated logit model: formulation and application to residential choice modeling. *Transportation Research Part B*, 38(2), 147-168.

Bhat, C.R., and J.Y. Guo (2007) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B,* 41(5), 506-526.

Bhat, C.R., and R. Sardesai (2006). The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B,* 40(9), 709-730.

Bhat, C.R., and I.N. Sener (2009) A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Journal of Geographical Systems*, 11(3), 243-272.

Bhat, C.R., and S. Srinivasan (2005) A multidimensional mixed ordered-response model for analyzing weekend activity participation. *Transportation Research Part B,* 39(3), 255-278.

Bhat, C.R., N. Eluru, and R.B. Copperman (2008). Flexible model structures for discrete choice analysis. In *Handbook of Transport Modelling*, *2nd edition*, Hensher, D.A., Button, K.J., (eds.), Elsevier Science, 75-104.

Bhat, C.R., I.N. Sener, and N. Eluru (2010a) A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B,* 44(8-9), 903-921.

Bhat, C.R., C. Varin, and N. Ferdous (2010b). A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered response model. In: Greene, W., Hill, R.C. (eds.), *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, Emerald Group Publishing Limited, 65-106.

Birnbaum, Z.W. (1950) Effect of linear truncation on a multinormal population. *Annals of Mathematical Statistics* 21(2), 272-279.

Birol, E., K. Karousakis, and P. Koundouri (2006) Using economic valuation techniques to inform water resources management: A survey and critical appraisal of available

techniques and an application. *Science of the Total Environment*, 365(1-3), 105-122.

Black, C., A. Collins, and M. Snell (2001) Encouraging walking: the case of journey-to-school trips in compact urban areas. *Urban Studies*, 38(7), 1121-1141.

Boarnet, M.G., C.L. Anderson, C. Day, T.E. McMillan, and M. Alfonzo (2005) Evaluation of the California safe routes to school legislation: urban form changes and children's active transport to school. *American Journal of Preventive Medicine*, 28 (suppl 2), S134-S140.

Bockstael, N.E. (1996) Modeling economics and ecology: the importance of a spatial perspective. *American Journal of Agricultural Economics*, 78(5), 1168-1180.

Bolduc, D., B. Fortin, and M. Fournier (1996) The effect of incentive policies on the practice location of doctors: a multinomial probit analysis. *Journal of Labor Economics,* 14(4), 703-732.

Brady, M., and E. Irwin (2011) Accounting or spatial effects in economic models of land use: recent developments and challenges ahead. *Environmental Resource Economics,* 48(3), 487-509.

Campbell, D., E. Doherty, S. Hynes, and T. Van Rensburg (2010) Combining discrete and continuous mixing approaches to accommodate heterogeneity in price sensitivities in environmental choice analysis. *84th Agricultural Economics Society Annual Conference*, March 29-31, Edinburgh, Scotland.

Capozza, D.R., and Y. Li (1994) The intensity and timing of investment: the case of land. *American Economic Review*, 84(4), 889-904.

Caragea, P.C., and R.L. Smith (2007) Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models. *Journal of Multivariate Analysis*, 98(7), 1417- 1440.

Carrión-Flores, C., and E.G. Irwin (2004) Determinants of residential land-use conversion and sprawl at the rural-urban fringe. *American Journal of Agricultural Economics,* 86(4), 889-904.

Carrión-Flores, C.E., A. Flores-Lagunes, and L. Guci (2009) Land use change: a spatial multinomial choice analysis. Presented at the Agricultural and Applied Economics Association 2009 Annual Meeting, July 26-28, 2009, Milwaukee, WI.

Cedilnik. A., K. Kosmelj, and A. Blejec (2006) Ratio of two random variables: a note on the existence of its moments. *Metodološki Zvezki - Advances in Methodology and Statistics,* 3(1), 1-7.

Chakir, R., and O. Parent (2009) Determinants of land use changes: a spatial multinomial probit approach. *Papers in Regional Science*, 88(2), 327–344

Chen, Y., S. Ravulaparthy, K. Deutsch, P. Dalal, S. Y. Yoon, T. Lei, K. G. Goulias, R. M. Pendyala, and C. R. Bhat (2010) Development of opportunity-based accessibility indicators. *Transportation Research Record,* 2255, 58-68

Cherchi, E., C. Cirillo, and J. Polak (2009) User benefit assessment in presence of random taste heterogeneity: comparison between parametric and nonparametric models. *Transportation Research Record,* 2132, 78-86.

Chintagunta, P.K., D.C. Jain, and N.J. Vilcassim (1991) Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research*, 28(4), 417-428.

Cho, W.T., and T. Rudolph (2007) Emanating political participation: untangling the spatial structure behind participation. *British Journal of Political Science* 38(2), 273-289.

City of Austin (2008) Neighborhood Planning: Guide to Zoning,

http://www.ci.austin.tx.us/planning/neighborhood/downloads/zoning_guide.pdf

Cooper, A.R., A.S. Page, L.J. Foster, and D. Qahwaji (2003) Commuting to school: are children who walk more physically active? *American Journal of Preventive Medicine*, 25, 273-276.

Cox, D.R., and N. Reid (2004) A note on pseudolikelihood constructed from marginal densities. *Biometrika*, 91(3), 729-737.

Daganzo, C., (1979) *Multinomial Probit: The Theory and its Application to Demand Forecasting*. Academic Press, New York.

Daly, A., S. Hess, and K. Train (2011) Assuring finite moments for willingness to pay in random coefficient models. *Transportation*, 39(1), 19-31.

Dellinger, A.M. and C.E. Staunton (2002) Barriers to children walking and biking to school – United States, 1999. *Journal of the American Medical Association*, 288(11), 1343-1344.

DiGuiseppi, C., I. Roberts, and L. Li (1998) Determinants of car travel on daily journeys to school: cross sectional survey of primary school children. *British Medical Journal*, 316, 1426–1428.

Dubin, R.A. (1998) Spatial autocorrelation: a primer. *Journal of Housing Economics*, 7(4), 304-327.

Elhorst, J.P. (2010a) Spatial panel data models. In: Fischer, M.M., Getis, A. (eds.), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, Springer, Berlin, 377-408.

Elhorst, J.P. (2010b) Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, 5(1), 9-28.

Ellison, B.E. (1964) Two theorems for inferences about the normal distribution with applications in acceptance sampling. *Journal of the American Statistical Association*, 59(305), 89-95.

Engle, R.F., N. Shephard, and K. Sheppard (2007) Fitting and testing vast dimensional time-varying covariance models. Finance Working Papers, FIN-07-046, Stern School of Business, New York University.

Engler, D.A., G. Mohapatra, D.N. Louis, and R.A. Betensky (2006) A pseudolikelihood approach for simultaneous analysis of array comparative genomic hybridizations. *Biostatistics* 7(3), 399-421.

Ewing, R., W. Schroer, and W. Greene (2004) School location and student travel: analysis of factors affecting mode choice. *Transportation Research Record*, 1895, 55-63.

Fleming, M.M. (2004) Techniques for estimating spatially dependent discrete choice models. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, Berlin, 145-168.

Fosgerau, M. (2005) Unit income elasticity of the value of travel time savings. Presented at 8th NECTAR Conference, Las Palmas G.C., June 2-4.

Fosgerau, M. (2006) Investigating the distribution of the value of travel time savings. *Transportation Research Part B*, 40(8), 688-707.

Franzese, R.J., and J.C. Hays (2008) Empirical models of spatial interdependence. In: Box-Steffensmeier, J.M., H.E. Brady, and D. Collier (eds.), *The Oxford Handbook of Political Methodology*, Oxford University Press, Oxford, 570-604.

Franzese, R.J., J.C. Hays, and L.M. Schaffer (2010) Spatial, temporal, and spatiotemporal autoregressive probit models of binary outcomes: estimation, interpretation, and presentation. APSA 2010 Annual Meeting Paper. Available at: http://ssrn.com/abstract=1643867

Genz, A. (1992) Numerical Computation of Multivariate Normal Probabilities, *Journal of Computational and Graphical Statistics*, 1, 141–149.

Genz, A. (2004) Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing*, 14(3), 251–260.

Glaeser, E.L., J. Gyourko, and R.E. Saks (2006) Urban growth and housing supply. *Journal of Economic Geography*, 6(1), 71-89.

Godambe, V.P. (1960) An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, 31(4), 1208-1211.

González-Farías, G., A. Domínguez-Molina, and A.K. Gupta (2004) Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, 126(2), 521-534.

Greene, W.H., and D.A. Hensher (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B*, 37(8), 681-698.

Greene W.H, D.A. Hensher, and J.M. Rose (2006) Accounting for heterogeneity in the variance of the unobserved effects in mixed logit models (NW transport study data). *Transportation Research Part B*, 40(1), 75-92.

Gupta, A.K., G. González-Farías, and A. Domínguez-Molina (2004) A multivariate skew normal distribution. *Journal of Multivariate Analysis*, 89(1), 181-190.

Hausman, J.A., and D.A. Wise (1978) A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica*, 46(2), 403-426.

Heagerty, P., and T. Lumley (2000) Window subsampling of estimating functions with application to regression models. *Journal of the American Statistical Association*, 95(449), 197-211.

Hensher, D.A., J.M. Rose, and W.H. Greene (2005) *Applied Choice Analysis: A Primer*. Cambridge University Press, Cambridge, U.K.

Hess, S., M. Bierlaire, and J.W. Polak (2007) A systematic comparison of continuous and discrete mixture models. *European Transport*, 37, 35-61

Hess, S., and J.M. Rose (2009). Allowing for intra-respondent variations in coefficients estimated on repeated choice data. *Transportation Research Part B,* 43(6), 708-719.

Hynes, S., N. Hanley, and R. Scarpa (2008) Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. *American Journal of Agricultural Economics,* 90(4), 1011-1027.

Irwin, E.G. (2010) New directions for urban economic models of land use change: incorporating spatial dynamics and heterogeneity. *Journal of Regional Science*, 50(1), 65-91.

Irwin, E.G. and N.E. Bockstael (2002) Interacting agents, spatial externalities, and the endogenous evolution of residential land use pattern *Journal of Economic Geography,* 2(1), 31-54.

Irwin, E.G., and J. Geoghegan (2001) Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment* 85(1-3), 7-24.

Jara, A., F. Quintana, and E. San Martín (2008) Linear mixed models with skew-elliptical distributions: a Bayesian approach. *Computational Statistics and Data Analysis*, 52(11), 5033-5045.

Joe, H. (1995) Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association,* 90(431), 957-964.

Jones, K., and N. Bullen (1994) Contextual models of house prices: a comparison of fixed- and random-coefficient models developed by expansion. *Economic Geography,* 70(3), 252-272.

Kamakura, W.A., and G.J. Russell (1989) A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research,* 26(4) 379-390.

Klier, T., abd D.P. McMillen (2008) Clustering of auto supplier plants in the U.S.: GMM spatial logit for large samples. *ASA Journal of Business & Economic Statistics*, 26(4), 460-471.

Koplan, J.P., C.T. Liverman, and V.I. Kraak (eds.) (2005) *Preventing Childhood Obesity: Health in The Balance*. Institute of Medicine of the National Academies, The National Academies Press, Washington, DC.

Lancaster, K. (1971) *Consumer Demand: A New Approach*. Columbia University Press, New York

Lawrence Frank and Company, Inc. (2008) Youth Travel to School: Community Design Relationships with Mode Choice, Vehicle Emissions, and Healthy Body Weight. Final Report, U.S. Environmental Protection Agency, Washington, D.C.

LeSage, J.P. (2000) Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* 32(1), 19-35.

LeSage, J.P., and R.K. Pace (2009) *Introduction to Spatial Econometrics*. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton.

Li, X., and X.P. Liu (2007) Defining agents' behaviors to simulate complex residential development using multicriteria evaluation. *Journal of Environmental Management*, 85(4), 1063-1075.

Li, Z., D.A. Hensher, and J.M. Rose (2010) Willingness to pay for reliability in passenger transport: a review and some new empirical evidence. *Transportation Research Part E*, 46(3), 384-403.

Lindsay, B.G. (1988) Composite likelihood methods. *Contemporary Mathematics*, 80(1), 221–240.

Loucaides, C. and R. Jago (2008) Differences in physical activity by gender, weight status and travel mode to school in Cypriot children. *Preventive Medicine*, 47, 107-111

Luce, R., and P. Suppes (1965) Preference, utility, and subjective probability. In *Handbook of Mathematical Psychology*, *Volume III*, Luce, R., R. Bush, and E. Galanter (eds.), John Wiley & Sons, New York.

McCulloch, R.E., and P.E. Rossi (2000) Bayesian analysis of the multinomial probit model. In *Simulation-Based Inference in Econometrics*, Mariano, R., Schuermann, T., Weeks, M.J., (eds.), 158-178, Cambridge University Press, New York.

McFadden, D. (1974) The measurement of urban travel demand. *Journal of Public Economics*, 3(4), 303-328.

McFadden, D. (1978) Modeling the choice of residential location. *Transportation Research Record* 672, 72-77.

McFadden, D., and K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447-470.

McDonald, N.C. (2007) Active transportation to school: trends among U.S. schoolchildren, 1969–2001. *American Journal Preventive Medicine*, 32(6), 509-516.

McDonald, N.C. (2008) Children's mode choice for school trip: the role of distance and school location in walking to school. *Transportation*, 35(1), 23-35.

McMillan, T.E. (2007) The relative influence of urban form on a child's travel mode to school. *Transportation Research Part A*, 41(1), 69-79.

McMillan, T.E., K. Day, M. Boarnet, M. Alfonzo, and C. Anderson (2006) Johnny walks to school does Jane? Sex differences in children's active travel to school. *Children, Youth, and Environments*, 16(1), 75-90.

McMillen, D.P. (2010) Issues in spatial analysis, *Journal of Regional Science*, 50(1), 119-141.

Meintanis, S.G., and Z. Hlávka (2010) Goodness-of-fit tests for bivariate and multivariate skew-normal distributions. *Scandinavian Journal of Statistics*, 37(4), 701-714.

Miller, H.J. (1999) Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geographical Analysis*, 31(4), 373-399.

Mitra, R., R.N. Buliung, and G.E.J. Faulkner (2010) Spatial clustering and the temporal mobility of walking school trips in the Greater Toronto Area, Canada. *Health & Place*, 16(4), 646-655.

Molenaar, D., C.V. Dolan, and N.D. Verhelst (2010) Testing and modeling non-normality with the one-factor model. *British Journal of Mathematical and Statistical Psychology*, 63(2), 293-317.

Molenberghs, G., and G. Verbeke (2005) *Models for Discrete Longitudinal Data*. Springer Series in Statistics, Springer Science + Business Media, Inc., New York.

Müller, G., and C. Czado (2005) An autoregressive ordered probit model with application to high frequency financial data. *Journal of Computational and Graphical Statistics*, 14(2), 320-338.

Muller, S., S. Tscharaktschiew, and K. Haase (2008) Travel-to-school mode choice modelling and patterns of school choice in urban areas. *Journal of Transport Geography*, 16(5), 342-357.

O'Hagan, A., and T. Leonard (1976) Bayes estimation subject to uncertainty about parameter constraints. *Biometrika*, 63(1), 201-203.

Pace, L., A. Salvan, and N. Sartori (2011) Adjusting composite likelihood ratio statistics. *Statistica Sinica,* 21(1), 129-148.

Pinkse, J., and M.E. Slade (1998) Contracting in space: an application of spatial statistics to discrete-choice models. *Journal of Econometrics,* 85(1), 125-154.

Plantinga, A.J., and E.G. Irwin (2006) Overview of empirical methods. In: Bell, K.P., Boyle, K.J., Rubin, J. (eds), *Economics of Rural Land-Use Change*, Aldershot, Ashgate Publishing, 113-134.

Pont, K., J. Ziviani, D. Wadley, S. Bennett, and R. Abbott (2009) Environmental correlates of children's active transportation: a systematic literature review. *Health & Place*, 15(3), 849-862.

Pooley, C., D. Whyatt, M. Walker, G. Davies, P. Coulton, and W. Bamford (2010) Understanding the school journey: integrating data on travel and environment. *Environment and Planning A*, 42(4), 948-965.

Sandor, Z., and K. Train (2004). Quasi-random simulation of discrete choice models. *Transportation Research Part B*, 38(4), 313-327.

Sener, I.N., and C.R. Bhat (2012) Flexible spatial dependence structures for unordered multinomial choice models: Formulation and application to teenagers' activity participation. *Transportation*, 39(3), 657-683

Sivakumar, A., C.R. Bhat, and G. Ökten (2005). Simulation estimation of mixed discrete choice models with the use of randomized quasi-Monte Carlo sequences: a comparative study. *Transportation Research Record,* 1921, 112-122.

Small, K.A., C. Winston, and J. Yan (2005) Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, 73(4), 1367-1382.

Smirnov, O.A. (2010) Spatial econometrics approach to integration of behavioral biases in travel demand analysis. *Transportation Research Record*, 2157, 1-10.

Timperio, A., K. Ball, J. Salmon, R. Robers, B. Giles-Corti, D. Simmons, L.A. Baur, and D. Crawford (2006) Personal, family, social, and environmental correlates of active commuting to school. *American Journal of Preventive Medicine*, 30(1), 45-51.

Torres, C., N. Hanley, and A. Riera (2011) How wrong can you be? Implications of incorrect utility function specification for welfare measurement in choice experiments. *Journal of Environmental Economics and Management,* 62(1), 111-121.

Train, K.E. (2003) *Discrete Choice Methods with Simulation*, 1st ed. Cambridge University Press, Cambridge.

Train, K.E. (2008) EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, 1(1), 40-69

Train, K.E (2009) *Discrete Choice Methods with Simulation*, 2nd ed., Cambridge University Press, Cambridge.

Train, K.E., and G. Sonnier (2005) Mixed logit with bounded distributions of correlated partworths. In *Applications of Simulation Methods in Environmental and Resource Economics*, Scarpa, R., Alberini, A., (eds.), Ch. 7, pp. 117-134, Springer, Dordrecht, The Netherlands.

Ulfarsson, G. and V. N. Shankar (2008) Children's travel to school: discrete choice modeling of correlated motorized and nonmotorized transportation modes using covariance heterogeneity. *Environment and Planning B*, 35(2), 195-206.

Varin, C., and C. Czado (2008) A mixed probit model for the analysis of pain severity diaries. Available at: http://www-m4.ma.tum.de/Papers/Czado/varin_czado_pain_diaries.pdf.

Varin, C., and P. Vidoni (2005) A note on composite likelihood inference and model selection. *Biometrika*, 92(3), 519-528.

Varin, C., and P. Vidoni (2009) Pairwise likelihood inference for general state space models. *Econometric Reviews*, 28(1-3), 170-185.

Varin, C., N. Reid, and D. Firth (2011) An overview of composite likelihood methods. *Statistica Sinica*, 21(1), 5-42.

Ver Hoef, J.M., and J.K. Jansen (2007) Space–time zero-inflated count models of harbor seals. *Environmetrics*, 18(7), 697-712.

Weinstein, M.A. (1964) The sum of values from a normal and a truncated normal distribution. *Technometrics*, 6(1), 104-105.

Yai, T., S. Iwakura, and S. Morichi (1997) Multinomial probit with structured covariance for route choice behavior. *Transportation Research Part B*, 31(3), 195-207.

Yang, S. and G.M. Allenby (2003) Modeling interdependent consumer preferences. *Journal of Market Research*, 40(3), 282-294.

Yarlagadda, A.K. and S. Srinivasan (2008) Modeling children's school travel mode and parental escort decisions. *Transportation*, 35(2), 201-218.

Yeung, J., S. Wearing, and A.P. Hills (2008) Child transport practices and perceived barriers in active commuting to school. *Transportation Research Part A,* 42(6), 895-900.

Zhao, Y., and H. Joe (2005) Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics,* 33(3), 335-356.

Zwerts, E., G. Allaert, D. Janssens, G. Wets, and F. Witlox (2009) How children view their travel behaviour: a case study from Flanders (Belgium). *Journal of Transport Geography,* 18(6), 702-710.